Online boiler convective heat exchanger monitoring: a comparison of soft sensing and data-driven approaches

Prepared by:

Gerto Prinsloo

PRNGER003

Department of Mechanical Engineering

University of Cape Town

Supervisors:

Professor Pieter Rousseau and Priyesh Gosai

November 2018

Submitted to the Department of Mechanical Engineering at the University of Cape Town in partial fulfilment of the academic requirements for a Master of Science degree in Mechanical Engineering

Key Words: coal fired power plant boiler, heat exchanger network, online monitoring, soft sensor, data-driven fault diagnosis

Abstract

Online monitoring supports plant reliability and performance management by providing real time information about the condition of equipment. However, the intricate geometries and harsh operating environment of coal fired power plant boilers inhibit the ability to do online measurements of all process related variables. A low-cost alternative lies in the possibility of using knowledge about boiler operation to extract information about its condition from standard online process measurements. This approach is evaluated with the aim of enhancing online condition monitoring of a boiler's convective pass heat exchanger network by respectively using a soft sensor and a data-driven method.

The soft sensor approach is based on a one-dimensional thermofluid process model which takes measurements as inputs and calculates unmeasured variables as outputs. The model is calibrated based on design information. The data-driven method is one developed specifically in this study to identify unique fault signatures in measurement data to detect and quantify changes in unmeasured variables. The fault signatures are initially constructed using the calibrated one-dimensional thermofluid process model. The benefits and limitations of these methods are compared at the hand of a case study boiler. The case study boiler has five convective heat exchanger stages, each composed of four separate legs. The data-driven method estimates the average conduction thermal resistance of individual heat exchanger legs and the flue gas temperature at the inlet to the convective pass. In addition to this, the soft sensor estimates the average fluid variables for individual legs throughout the convective pass and therefore provides information better suited for condition prognosis.

The methods are tested using real plant measurements recorded during a period which contained load changes and on-load heat exchanger cleaning events. The cleaning event provides some basis for validating the results because the qualitative changes of some unmeasured monitored variables expected during this event are known. The relative changes detected by both methods are closely correlated. The data-driven method is computationally less expensive and easily implementable across different software platforms once the fault signatures have been obtained. Fault signatures are easily trainable once the model has been developed. The soft sensors require the continuous use of the modelling software and will therefore be subject to licencing constraints.

Both methods offer the possibility to enhance the monitoring resolution of modern boilers without the need to install any additional measurements. Implementation of these monitoring frameworks can provide a simple and low-cost contribution to optimized boiler performance and reliability management.

Declaration

I, Gerto Prinsloo, hereby declare the work contained in this dissertation to be my own. All information which has been gained from various journal articles, text books or other sources has been referenced accordingly. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work or part thereof.

Signed by candidate

Gert Thomas Prinsloo

2018/10/21

Date

Acknowledgements

I acknowledge that my capability to observe and reason was created through the First and the Last, who died and came to life, to who I am thankful for offering me life in Him.

To my wife, my loving companion throughout this journey, I will always be grateful.

I am grateful to my supervisors, Professor Pieter Rousseau and Priyesh Gosai, not only for imparting their plentiful knowledge, but also for their mastery in the art of guidance, through which this research exceeded at least my own expectations.

Associate Professor Wim Fuls at the University of Cape Town (UCT) contributed the methodology presented in Section 4.2.3.

M-Tech Industrial (Pty) Ltd is acknowledged for allowing the use of Flownex®.

This research was produced by the initiative of the Eskom Power Plant Engineering Institute (EPPEI) in collaboration with the Applied Thermofluid Process Modelling (ATProM) research unit at the University of Cape Town (UCT) and funded by Eskom.

Table of Contents

List of Figures

List of Tables

List of Nomenclature

General symbols

A	Area
c	Specific heat
C	Heat capacity
C_r	Ratio of the minimum to the maximum heat capacity of two fluids
D_H	Hydraulic diameter
f	Darcy-Weisbach friction factor
F	Correction factor for the log mean temperature difference
h	Enthalpy
H	Constant for the Hilpert correlation
k	Thermal conductivity
K	Pressure loss coefficient
L	Length
m	Constant applied in the Hilpert correlation
$\dot{m}$	Mass flow rate
M	Molar mass
NTU	Number of transfer units
Nu	Nusselt number
p	Pressure
Pr	Prandtl number
$\dot{Q}$	Heat transfer rate
r	Radius
R	Thermal resistance
R_g	Specific gas constant
$\overline{R}_g$	Universal gas constant
Re	Reynolds number
SPE	Squared prediction error
T	Temperature
TS	Hotelling's T-squared statistic
u	Heat transfer coefficient
U	Overall heat transfer coefficient
v	Scalar quantity of a variable
w	Velocity
x	Mass fraction

Greek symbols

α	Confidence interval
δ	Threshold on the squared prediction error index
Δ	Difference
ε	Heat transfer effectiveness
ε_r	Emissivity of an object
ε_s	Surface roughness
μ	Viscosity
ρ	Density
σ	Stefan Boltzmann constant
φ	Combined index
χ	Threshold on the Hotelling's T-squared index

Vectors and matrices

f	Fault vector
F	Fault matrix
i	Fault index
Λ	Diagonal matrix of ordered eigenvalues of a covariance matrix
Ξ	Fault library
Σ	Covariance matrix
T	Principal component scores
U	Principal component loading vectors
x	Vector of a single sample of measurements
X	Normalised matrix of measurements
V	Right singular vectors of a covariance matrix

Acronyms and Abbreviations

ABC	Angle-based contribution
ANN	Artificial neural network
BMCR	Boiler maximum continuous rating
CFPP	Coal fired power plant
DCS	Distributed control system
EPPEI	Eskom Power Plant Engineering Institute
EPRI	Electric Power Research Institute
EWMA	Exponentially weighted moving average
HE	Heat exchanger
IAPWS	International Association for the Properties of Water and Steam
LMTD	Log mean temperature difference
NDE	Non-destructive evaluation
NIST	National Institute of Standards and Technology
NOP	Normal operating point
NOR	Normal operating region
PCA	Principal components analysis
PC	Principal component
PCS	Principal component subspace
RBC	Reconstruction based contribution
RH	Reheater
RS	Residual subspace
SH	Superheater
TLD	Tube leak detector

1. Introduction

1.1 Background and motivation

Contemporary society strives for continuous economic growth, which is fuelled by the availability of electrical energy. Globally, 40% of electricity consumed is generated using coal fired power plants (CFPP) (International Energy Agency 2016). CFPPs, owned and operated by Eskom, generate more than 91% of the electricity consumed in South Africa (Eskom 2017). Optimal management of CFPP equipment reliability and performance is the key to economic electricity supply worldwide.

Superheated steam, used in the core cycle of these plants, is generated using boilers. These boilers consist of several heat exchanger stages arranged in a network of steam carrying tubes exposed to heat released from coal combustion. Being critical equipment, boiler failures lead to plant down time, resulting in a shortfall in generation capacity, cost increases due to production loss and repairs, increased safety risks and the premature exhaustion of plant life because of additional shut-downs and start-ups. Despite boiler reliability management therefore being amongst the highest priorities for power plant owners, boiler failures remain the single largest contributor to unplanned downtime in these plants (Coleman 2011a).

The Electric Power Research Institute (EPRI) has done valuable research on the reduction of boiler failures and consolidated a report that provides an industry wide approach towards the management of boiler reliability (Coleman 2011a). The generic technical approach throughout their report is facilitated by offline monitoring to assess boiler condition by means of non-destructive examination (NDE), while online monitoring of process measurements is used to ensure that selected boiler design limits are not exceeded.

Because of the large surface area of boiler heat exchangers, tight outage schedules, limited resources and access restrictions, NDE inspections are not always exhaustive. Changes in operating conditions change the distribution and rates of damage, further impeding failure prediction and prevention. Consequently, boiler reliability remains a difficult and expensive challenge.

The state of the art in reliability management is to predict failures (Hoppenstedt et al. 2018). For a given knowledge base, prognosis improvement lies in increasing the extent and rate of monitoring. Real time monitoring during boiler operation provides such an opportunity. However, NDE inspections during boiler operation are not feasible with the current measurement technologies, mostly because of the harsh flue gas environment and intricate boiler geometries.

Research into the development of new online measurement technologies has not yet yielded an industrially practicable and cost-effective solution (Pena et al. 2013).

Boiler tubes ultimately fail because of stress raisers like discontinuities and wall thinning. Stress raisers are caused by specific damage mechanisms. Most of these damage mechanisms are inevitably present in CFPP boilers and are allowed for in the design. However, several factors, like design assumptions or off-design operation, cause these damage mechanisms to be amplified in sections of the boiler leading to premature failure (Coleman 2011a). Some damage mechanisms are driven by fluid and heat exchanger variables, while changes in some of these variables are in turn driven by damage mechanisms. Online measurement of the fluid and heat exchanger variables are however also inhibited by the high temperature and abrasiveness of the flue gas and its intricate flow path geometry. But even though these fluid and heat exchanger variables are not directly measured, they have a significant impact on standard online boiler process measurements of flow and temperature. The location and severity of damage mechanisms are therefore potentially observable during boiler operation through the effects that are observed in existing online process measurements. What makes this approach particularly attractive is that there is no need to install additional measurements. This means that it can potentially be implemented at relatively low cost.

Much development has occurred in online monitoring and diagnostic capabilities and these have been successfully implemented on a wide range of industrial applications (Tidriri et al. 2016). The power industry is therefore challenged to research the possibilities of improving boiler reliability and performance management through the application of online monitoring and diagnostic techniques.

This study aims to support the current boiler reliability improvement program by using standard online process measurements to pro-actively identify circumstances that exacerbate deterioration of boiler heat exchanger condition and performance during operation. This presents the prospect of optimizing operational practices in terms of boiler reliability or performance and having an additional tool for informing maintenance decisions and design changes.

1.2 Research problem

Given that standard boiler process measurements of temperature and flow may contain more information about the condition of a boiler than is immediately evident, the research problem is one of extracting this additional information from the measurements. This requires a means of relating the behaviour of measurements to the state of unmeasured variables. This can be done based on a theoretical understanding of the process or from directly observing these interactions.

Using the values of measured variables as inputs, the values of certain unmeasured variables may be calculated by solving the equations that describe the theoretical understanding of the thermofluid processes in the boiler. This approach provides the absolute values of these unmeasured variables and will be referred to as a soft sensor. A soft sensor is therefore a process model devoted to the estimation of unmeasured plant variables.

Previous observations of process interactions can also be used to match patterns in observed data to information about unmeasured variables. This approach is therefore termed data-driven. The outputs of data-driven methods do not necessarily provide the absolute values of process variables but rather allow conclusions to be drawn about the presence, severity and causes of deviations from normal operation. The use of data-driven methods for monitoring is therefore approached from a fault detection and diagnosis perspective.

A soft sensor based on fundamental thermofluid relations is limited in its application to real time monitoring because of its high computational and specialised software requirements. Data-driven methods provide a promising alternative which is computationally cheaper since it does not require continuous simulation using specialised software during online monitoring. It is also thought that the patterns in measured data might contain different or additional information about unmeasured variables than that obtained with soft sensors based on the same measurements. However, whereas the accuracy of the soft sensor is determined by the accuracy of the process model, the quality of data-driven results hinges on the quality of its training data. Obtaining high quality training data for boilers presents several practical limitations. The greatest being that the variables of interest are neither measured nor controlled. Training data can however be produced by simulation using a thermofluid process model. This approach has the benefit that different fault types for all operational modes can readily be emulated in the model. This provides the possibility of training a data-driven method with capabilities nearing that of the soft sensor itself.

A model can therefore be employed in online monitoring either by its direct use as a soft sensor, or its indirect use by providing training data for a data-driven method. To develop an optimal online boiler monitoring framework, these two approaches must be compared with respect to their relevant capabilities, benefits and limitations.

1.3 Objectives of this study

The primary objective of this study is to investigate and compare the potential of applying a soft sensor and a data-driven approach in online monitoring of boiler heat exchangers using only standard measured process parameters as inputs.

The enabling objectives are to:

I. Develop a simplified thermofluid process model of a case study boiler.
II. Validate/calibrate the thermofluid process model using design data of the real plant.
III. Apply this model as a soft sensor.
IV. Select or derive a data-driven monitoring method applicable to boiler monitoring.
V. Train the data-driven method using the developed process model.
VI. Compare the capabilities, benefits and limitations of the soft sensor and the data-driven method at the hand of recorded real plant process data.

1.4 Assumptions and scope of this study

The scope of this study includes an evaluation and discussion of the monitoring results of real plant data from a case study boiler obtained with two methods, namely the direct and indirect use of a one-dimensional thermofluid process model respectively. However, the focus is not on developing highly accurate process models. The focus is on the methodologies used for employing these models in online monitoring.

Wall tubes in the convective pass, heat losses, directed radiation and ingress air are neglected in the model development. The heat exchangers considered in this study are limited to the convective heat exchanger stages carrying superheated steam and excludes the evaporator and the economiser.

This study is concerned with monitoring variables that impact the condition and performance of boilers and does not aim to identify faulty sensors and actuators.

2. Literature Review

2.1 Effects of boiler process variables on damage mechanisms

Some damage mechanisms in CFPP boilers are driven by process variables. The state and distribution of these process variables are therefore of great interest to boiler health prognosis. Tube erosion by fly ash entrained in the flue gas is amongst the most dominant damage mechanisms in boilers (Dooley & Chang 2000). Fly ash erosion is proportional to the ash particle velocity to powers ranging between three and four depending on the ash particle characteristics (Mbabazi et al. 2004). The particle velocity, driven by the flue gas velocity, is therefore a key driver of fly ash erosion damage. Flue gas flow maldistribution consequentially leads to nonuniform tube wear rates (Coleman 2011b). Flue gas flow maldistribution can be caused by localised fouling in gas passages or nonuniform heat distributions. Fly ash erosion rates are also influenced by the temperature of both ash particles and the tube metal. Ash particles are more abrasive at lower temperatures (Coleman 2011b), while the greater ductility of tube metal at elevated temperatures lead to accelerated erosion damage (Das et al. 2006).

Microstructural degradation of tube materials due to metal operating temperatures exceeding the design values are another major cause of boiler failures (Dooley & Chang 2000). Nonuniform pressure drops and flow distributions on either the steam or gas sides can substantially accelerate long term overheating damage (Coleman 2011a). Xu et al. (2000) has shown that major steam temperature differences can occur across the width of a heat exchanger stage and that the steam temperature distribution is indirectly proportional to the steam flow distribution. According to Trojan & Taler (2015), local steam flow rates decrease with increased local flue gas heating. Therefore, in addition to the higher steam temperatures caused by a higher temperature difference driving heat transfer, the steam flow rate drops, providing less cooling to the tube and consequently the tube metal temperatures are severely affected. Such variations in flow are functions of operating conditions and are accentuated during low load operation (Coleman 2011a). Increased tube metal temperatures accelerate the growth of the tube internal oxide layer. In turn, the high thermal resistance of oxide layers reduces the amount of tube cooling from the steam flow, leading to elevated tube metal temperatures (Cardoso et al. 2012). The oxide scale method for remaining life assessment of superheater and reheater tubes has become prevalent and is used for boilers worldwide (Coleman 2011a).

2.2 Online monitoring of boiler convective heat exchangers

Online monitoring improvement is either aimed at the development of measurement instrumentation, to get measurements where it was previously unfeasible, or at indirectly estimating unmeasured variables from measurements that are more feasible to implement. The literature reviewed in this section, in accordance with the objective of this study, covers the latter approach.

Research about online monitoring of convective heat exchangers in a CFPP boiler revolves mainly around estimating the severity of ash fouling on heat exchanger tubes through its effect on conduction heat transfer resistance. The goal of research in this field is often focused on cleaning practices for optimized performance.

While heat flux measurements are the preferred method for monitoring slagging on furnace walls during operation, its feasibility for application in convective pass heat exchangers is limited because of the principles of measurement, erosion by ash and heat transfer area requirements. Research on convective heat exchanger monitoring is therefore aimed at indirectly estimating parameters of heat transfer models related to fouling from measurements that are more feasible to implement online (Valero & Cortbs 1996).

Taler et al. (2009) suggests a fouling monitoring system based solely on steam side measurements for the superheat section of a non-reheat coal fired boiler. The superheat section consists of three superheater stages. Their approach estimates the actual rate of heat transfer to the lumped superheat section using an energy balance. The actual heat transfer rate is then baselined against that expected for clean tubes to give a heat absorption factor. They suggest that the heat absorption factor should provide an indication of fouling. The expected heat transfer rate used in their study is obtained experimentally as a function of load, although the details of the experimental setup are not given. This heat absorption factor is however also affected by process variables deviating from those present during the experimental acquisition of the clean heat transfer rates.

To disconnect the effects of fouling from those caused by changes in process properties using heat transfer models requires information about some flue gas process variables. Direct measurement of these variables is inhibited by the harsh environment and intricate geometries of the flue gas flow path. Standard boiler measurements generally allow the estimation of flue gas flow rates by balancing the flows entering the boiler (ASME 2008). An energy balance around the economiser can also produce the flue gas flow rate, given that the inlet and outlet flue gas temperatures are measured with the economiser heat load known (Shi et al. 2015). Availability of the flue gas flow rate allows an estimation of the interstage flue gas temperatures between heat exchanger stages

with known heat transfer rates if the temperature at any point in the convective pass is known (Cantrell & Idem 2010). This is however a crude approximation because of the difficulty of accurately accounting for radiation losses, directed radiation and ingress air.

Cantrell & Idem (2010) proposes a fouling monitoring index similar to that described by Taler et al. (2009). However, the heat transfer rate expected for clean tubes is obtained from online simulation of a flue gas side model which accounts for the effects of changing flue gas variables on tube external convective heat transfer. They also extend the monitoring resolution to each convective heat exchanger stage. This is made possible by the interstage steam temperature measurements from which the actual heat transfer rate to each stage can be determined.

The fouling monitoring system developed by Shi et al. (2015) calculates a cleanliness factor for different heat exchanger stages in the convective pass of a coal fired boiler. This cleanliness factor is the ratio of the actual overall heat transfer coefficient to that expected for clean tubes at the actual operating conditions. The actual overall heat transfer coefficient is obtained using the log-mean temperature difference (LMTD) (Incropera et al. 2007). The expected overall heat transfer coefficient for clean tubes in their study is calculated using correlations that account for tube external radiation and convection heat transfer. A similar approach is employed by Pattanayak et al. (2015) who only accounts for tube internal and external convection heat transfer when calculating the expected heat transfer coefficient.

Díez et al. (2005) proposes direct monitoring of the conduction thermal resistance by ash fouling. This is achieved by subtracting from the overall resistance that caused by external convection and radiation from the flue gas, conduction through the tube wall and internal convection to the steam. They incorporate a correction factor in the LMTD accounting for the specific configuration of each heat exchanger stage.

Data-driven approaches towards boiler monitoring are generally researched with the aim of boiler trip prevention and tube leak detection. This is achieved by analysis of patterns in data. These patterns are indicative of a high probability that conditions are either similar to that experienced at the onset of previous trips and leaks, or different from that expected for normal operation.

Alnaimi & Al-Kayiem (2011) claims that artificial neural networks (ANN) provide an artificial form of intelligence that allows diagnosis of low superheater temperatures. The ANN is trained on historical operating data. Later work extending on this study employs the same method to forecast boiler trips being caused by tube leaks, in which the achievement of a ten minute early notification period is reported (Ismail et al. 2016). The input layer of the ANN includes 26 measurements of various related process variables and excludes tube leak detector measurements.

Ma et al. (2014) develops an ANN for detecting tube leaks, which they refer to as four-tube leaks, in a coal fired boiler. The data used for training this ANN is obtained from simulating a boiler model for a 600 MW CFPP. The same boiler model is used to generate data for testing the algorithm. The minimum superheater leak size tested for is 8% of the rated evaporation rate, which represents a severe leak. From these tests the authors conclude that this method can successfully detect a leak under variable loads. The input layer of the ANN includes 20 measurements of various related process variables and excludes tube leak detector measurements.

The study by Rostek et al. (2015) aims to detect and isolate tube leaks in a fluidized bed boiler. Threshold evaluation of four individual leak sensitive variables is compared to leak detection and isolation using an ANN. It is reported that both methods can detect leaks in most locations at least two days before boiler shutdown is granted. The ANN classifier however detects some leaks that goes unnoticed by trending of individual variables, while being able to isolate the location of the leak as being either in the furnace, economiser or superheater sections. The input layer of the ANN includes 12 measurements of various related process variables and excludes tube leak detector measurements.

The successful use of principal components analysis (PCA), k-means clustering and Mahalanobis distance respectively to detect the presence of coal fired boiler tube leaks through deviation of measurement data from a normal operating region has been reported (Yu et al. 2015; Yu, Jang, Yoo, Park, et al. 2016; Yu, Jang, Yoo & Kim 2016). In these studies, the normal operating region is defined using historical data captured in the absence of a tube leak. The results are smoothed using an exponentially weighted moving average (EWMA) to reduce the number of false alarms. Measurements of 13 relevant process variables are considered, with the exclusion of tube leak detector measurements.

2.3 Online fault diagnosis

Online analysis of systems to establish its fault status holds numerous prognostic benefits and has been extensively researched. Fault diagnosis enables useful information about the current state of a process to be extracted and support the decision making process related to the control and optimization of operating actions (Jämsä-Jounela 2007) and maintenance planning.

Fault diagnosis generally consist of three tasks namely fault detection, fault isolation and fault identification (Frank et al. 2000). When a fault occurs in a system in such a way that its presence can become known by the monitoring framework, a fault is detected. Once a fault has been detected, it is desirable to isolate the actual fault from all possible faults. Fault isolation deals with

locating the position or the type of fault. Further analysis of the fault severity and root cause is termed fault identification. Fault identification enables corrective action to be taken to bring the system back to normal operation, or, if the problem cannot be corrected, to predict the effects of the out of normal operating condition on plant health and performance.

Online fault diagnosis methods are commonly grouped into two broad categories being model-based and data-driven methods.

Model-based methods leverage the analytical redundancy of a model by running the model next to the plant to generate residuals, which is the difference between measured process variables and their estimates (Frank 1990). Diagnostic observers (Clark 1979), parity relations (Gertler 1991), Kalman filters (Fathi et al. 1993) and parameter estimation (Isermann 1984) have been used for such residual generators. A fault is detected when this residual exceeds a specified threshold (Ding 2008). Diagnosis using model-based methods is achieved through evaluation of enhanced residuals. Enhanced residuals are subsets of residuals that selectively respond only to specific faults (Tidriri et al. 2016; Venkatasubramanian, Rengaswamy & Kavuri 2003). Generating enhanced residuals has been approached using structured residuals (Gertler & Singer 1990), directional residuals (Gertler & Monajemy 1995) and diagonal residuals (Gertler 1997).

Data-driven fault diagnosis methods are only dependent on the measured process variables and do not require a model of the process (Yin et al. 2012). These methods identify features in measurement data indicative of the fault status (Jämsä-Jounela 2007). Principal components analysis (PCA) (Hotelling 1933) is a popular data-driven fault diagnosis method with excellent fault detection capabilities in highly correlated multivariate datasets (MacGregor & Kourti 1995). A linear model for normal process operation is constructed using PCA and faults are detected by deviations from this model. Fault isolation based on PCA has been approached using contribution plots (MacGregor et al. 1994; Miller et al. 1998), reconstruction-based contribution (RBC) (Alcala & Qin 2009) and angle-based contribution (ABC) (Raich & Cinar 1996; Seongkyu Yoon & MacGregor 2001). Contribution plots identify variables which contribute most to a deviation from the normal operating model. It has been shown that contribution plots are inherently flawed and can misdiagnose even single faults that affect only one variable (Alcala & Qin 2009; Zhou et al. 2016). The RBC and ABC methods are essentially the same and will result in the same diagnosis (Alcala & Joe Qin 2011). The RBC method isolates the fault direction capable of reconstructing the faulty sample back into the normal operating region defined by the PCA model. Zhou et al. (2016) has shown that the RBC method can misdiagnose faults that affects more than one variable simultaneously and that the method is only suited for unidimensional faults. Despite many efforts to achieve fault isolation using PCA and extensions to this method, no complete solution has been achieved (Venkatasubramanian, Rengaswamy, Kavuri, et al. 2003). Artificial neural networks (ANN)

is a data-driven method with better fault isolation capabilities as it essentially classifies arbitrary regions in the data space (Cybenko 1989).

In the review by Katipamula & Brambley (2005) it is noted that the process of selecting an appropriate fault diagnostic method has not been generalized and is specific to the type of system, the faults anticipated and the sensor configuration. The authors contrast the higher accuracy, reliability, modelling complexity and computational intensity of model-based diagnostic methods to that of data-driven methods. The choice for a diagnostic system is therefore usually based on a trade-off between these characteristics.

In some cases, a combination of model-based and data-driven methods provides the most practical solution to the implementation of a fault diagnosis system. Such combinations are called hybrid methods (Venkatasubramanian, Rengaswamy, Kavuri, et al. 2003). Most hybrid methods use data-driven methods for fault detection, followed by the use of model-based methods for fault isolation (Tidriri et al. 2016). This is because of the lower computational requirements of data-driven techniques, providing quick detection, and the generally more descriptive nature of model-based methods. In most practical applications, the validation of results are challenging and the implementation of more than one fault diagnosis systems may generate conflicting results (Tidriri et al. 2016).

Process fault diagnosis methods were reviewed by Venkatasubramanian, Rengaswamy, Yin, et al. (2003) according to ten performance metrics. One metric deemed especially relevant for boiler monitoring is the ability to isolate multiple faults that occur simultaneously. This is because boiler operation is seldomly fault free and faults are rarely expected to occur in isolation or be individually present. Their review has found model-based diagnostic methods capable of diagnosing multiple faults occurring simultaneously, while no data-driven method was found with this capability. It is however possible to use data-driven methods for multiple simultaneous fault diagnosis if that specific combination of faults were incorporated in the training data. The authors note that this approach quickly becomes “combinatorially prohibitive” for large systems with many possibilities of different ratios of fault combinations. The joint effect of overlapping variables in interconnected processes makes multiple simultaneous fault diagnosis one of the great challenges for data-driven fault diagnosis (Chiang et al. 2015).

Another major challenge facing the fault diagnosis community is the problem of dealing with nonlinearities in process parameter interactions (Yin et al. 2014). These nonlinearities cause the response of process parameters incited by faults to change under different operating conditions and under the presence of other faults, which greatly increases the difficulty of designing an accurate diagnostic system.

Soft sensors are yet another approach less commonly acknowledged by the fault diagnosis research community. Soft sensors, also called inferential models or virtual sensors, are models devoted to the estimation of plant variables (Fortuna et al. 2007). These models can be either built from a physical understanding of the process (Li et al. 2011) or identified from historical process data (Si et al. 2009). Many research efforts has been focussed on establishing thresholds for data-driven fault detection indices (Qin 2003; Ricardo Dunia & Qin 1998; Alcala & Joe Qin 2011; Rato & Reis 2014). With soft sensors, the task of setting up detection thresholds is greatly simplified as the diagnosis results are presented in the form of actual parameter values (Haves et al. 1996).

2.4 Summary

In the literature reviewed, soft sensor approaches have been employed to monitor some unmeasured boiler variables related to damage mechanisms, while data-driven approaches are limited to the detection of existing tube leaks.

Soft sensor approaches based on thermofluid models have been employed to monitor the average conduction thermal resistance of respective heat exchanger stages in a boiler's convective pass, as well as the average flue gas temperatures between these stages. Although these models are developed with the aim of informing on load cleaning practices, these variables can also be used to inform predictions on boiler overheating and erosion damage.

Boiler damage is driven by process variables. Design deviations and maldistribution of several of these variables can occur simultaneously. Data-driven approaches experience difficulties when dealing with multiple simultaneous faults. An online monitoring framework employing a process model therefore allows more detailed information about boiler condition to be extracted from measurements.

3. Theory

3.1 Boiler convective pass thermofluid relations

3.1.1 Fluid properties

The National Institute of Standards and Technology (NIST) provides standardized reference data for the thermodynamic properties of water, steam and air for a range of thermodynamic states. The water and steam data provided by NIST are based on the standards of the International Association for the Properties of Water and Steam (IAPWS). These properties are however not standardized for flue gas as it depends on the coal composition and combustion characteristics. Of specific interest is the enthalpy and density of the flue gas in relation to its temperature and pressure. These properties directly impact energy and momentum conservation.

The flue gas enthalpy is composed by the enthalpies of its individual constituents and can therefore be determined if its composition is known as follows:

$$h_{fluegas} = \sum x_i h_i \tag{3.1}$$

with *h* the enthalpy and x_i the mass fraction of the i^{th} constituent. The enthalpies of the different constituents are available from NIST.

Using the ideal gas law, the flue gas density ρ at a given temperature *T* and pressure *p* is determined by the specific gas constant R_g.

$$\rho = \frac{p}{R_g T} \tag{3.2}$$

The specific gas constant for flue gas is composed of the specific gas constants of its constituents.

$$R_{g_{fluegas}} = \sum x_i R_{g_i} \tag{3.3}$$

The specific gas constant for each constituent can be obtained using the universal gas constant $\bar{R}_g$ and the molecular mass *M* of that substance.

$$R_g = \frac{\bar{R}_g}{M} = \frac{8.3144598}{M} \frac{J}{mol \cdot K} \tag{3.4}$$

3.1.2 Heat transfer

The heat transfer effectiveness of a heat exchanger, ε, is given by the ratio of the actual heat transfer rate, $\dot{Q}$, to the maximum possible heat transfer rate:

$$\varepsilon = \frac{\dot{Q}}{\dot{Q}_{\max}} \tag{3.5}$$

The maximum possible heat transfer is limited by the minimum heat capacity and the maximum temperature difference of the two fluids as follows:

$$\dot{Q}_{\max} = C_{\min}\left(T_{\max} - T_{\min}\right) \tag{3.6}$$

The heat capacity of a fluid is the heat input or output required for a specified amount of the fluid flow to undergo a temperature change:

$$C_{fluid} = \dot{m}_{fluid} c_{fluid} \tag{3.7}$$

with *c* the specific heat and $\dot{m}$ the mass flow rate.

The effectiveness of a heat exchanger is a function of the flow arrangement and heat exchanger layout and can be formulated using the number of transfer units, *NTU*, and the ratio of the minimum to the maximum heat capacity of the fluids between which heat is transferred, C_r (Holman 2010). A suitable correlation for the typical cross-flow arrangement of boiler heat exchangers has been empirically expressed as (Kays & London 1984):

$$\varepsilon = 1 - e^{\frac{NTU^{0.22}}{C_r}\left(e^{-C_r NTU^{0.78}} - 1\right)} \tag{3.8}$$

The *NTU* is a dimensionless number representing the ratio of the overall heat transfer coefficient, *UA*, to the minimum heat capacity:

$$NTU = \frac{UA}{C_{\min}} \tag{3.9}$$

With the actual heat transfer rate in equation (3.5) known, the *NTU* can be numerically determined with equation (3.8), allowing the overall thermal resistance of the heat exchanger to be established:

$$R_{overall} = \frac{1}{UA} \tag{3.10}$$

The overall thermal resistance also effects a total rate of heat transfer given a temperature difference across a heat exchanger such that it can be acquired from the relation (Incropera et al. 2007):

$$R_{overall} = \frac{F\Delta T_{LMTD}}{\dot{Q}} \tag{3.11}$$

with ΔT_{LMTD} the log mean temperature difference across the heat exchanger. The theoretical derivation of ΔT_{LMTD} is based on either pure parallel- or counter-flow heat exchanger orientations. The heat exchanger orientations found in CFPP boilers are typically multi-pass cross-flow and a correction factor *F* is therefore introduced to account for these deviations. The factor *F* is the ratio between the *NTUs* for pure counter flow and that flow arrangement under consideration. The *NTU* for pure a counter flow arrangement is given by (Kays & London 1984):

$$NTU = \begin{cases} \frac{1}{C_r - 1}\ln\left(\frac{\varepsilon - 1}{C_r\varepsilon - 1}\right) & \left(C_r < 1\right) \\ \frac{-\varepsilon}{\varepsilon - 1} & \left(C_r = 1\right) \end{cases} \tag{3.12}$$

The overall thermal resistance is composed of the sum of thermal resistances by convection and radiation from the flue gas to the tube external surface, conduction from the tube external surface through the ash, steel and oxide layers, and finally convection from the tube internal surface to the steam. The diagram in Figure 3-1 depicts these thermal resistances overlaid on a cross section of the tube and transferring mediums.

r_e

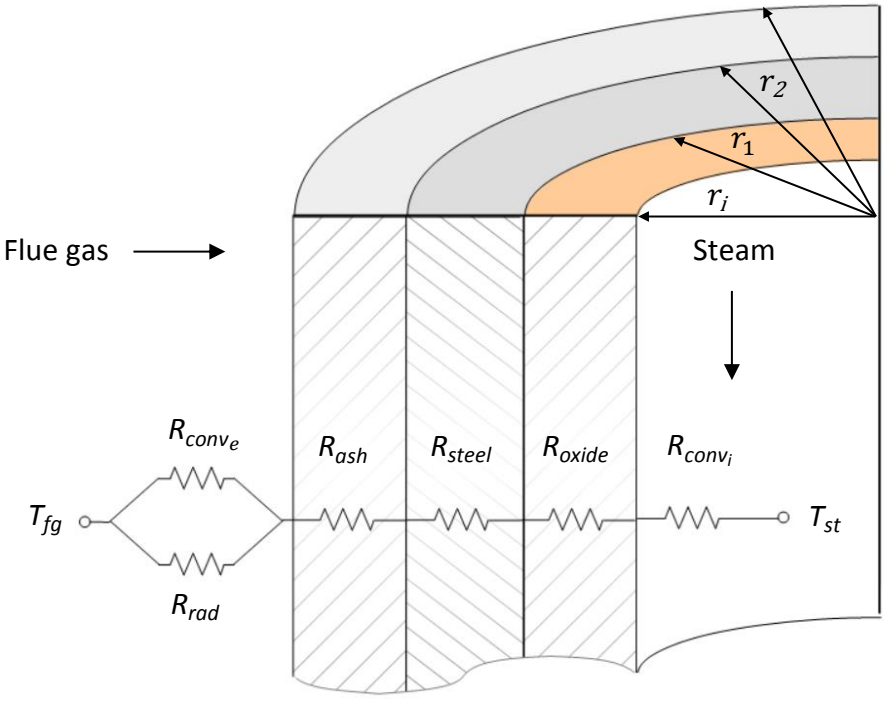


Figure 3-1: Tube cross section showing the components of thermal resistance

The value of each of these thermal resistances is in turn functions of the fluid variables and the geometry and composition of the heat exchanger.

The internal convection thermal resistance is orders of magnitude less than that of the external convection, radiation and conduction through fouling or oxide scale and is therefore often neglected.

The external convection thermal resistance is given by:

$$R_{conv_e} = \frac{1}{u_{conv_e} A_e} \tag{3.13}$$

with A_e the contact area between the mediums and u_{conv_o} the external convection heat transfer coefficient given by:

$$u_{conv_e} = \frac{Nu_{conv_e} k_{fg}}{L} \tag{3.14}$$

with k_{fg} the conductivity of flue gas and L the characteristic length as specified when developing the correlation for the non-dimensional Nusselt number, Nu.

The choice for the correlations used in this study is based on simplicity and widespread use. Although the absolute value of conduction thermal resistance bears potentially important information about the heat exchanger condition, greater focus is placed here on monitoring changes in this value relative to some baseline and correlating its development to operating

actions and time. The purpose of these correlations is therefore rather to account for the bulk effect that changing operating conditions have on different mechanisms of heat transfer. The external convection thermal resistance, R_{conv_e}, is calculated using the Hilpert correlation (Hilpert 1933) for the Nusselt number as follows:

$$Nu_{conv_e} = H\mathrm{Re}^m \mathrm{Pr}^{0.33} \tag{3.15}$$

The values of *H* and *m* vary according to the Reynolds number as shown in Table 3-1. Re and Pr are respectively the non-dimensional Reynolds and Prandtl numbers determined using the flue gas flow properties for each specific heat exchanger:

$$\mathrm{Re} = \frac{\rho w D_H}{\mu} \tag{3.16}$$

$$\mathrm{Pr} = \frac{c\mu}{k} \tag{3.17}$$

with μ the viscosity, *w* the fluid velocity and D_H the hydraulic diameter of the flow area.

Table 3-1: Variables for Hilpert correlation as a function of Reynolds number

Re	***H***	***m***
0.4 – 4	0.989	0.33
4 – 40	0.911	0.385
40 – 4000	0.683	0.466
4000 – 40 000	0.193	0.618
40 000 – 400 000	0.027	0.805

Heat is transferred between the flue gas and tube external surface by parallel mechanisms of convection and radiation. The temperature difference between the flue gas and the external tube surface effects a radiation heat transfer rate, which can be approximated by (Incropera et al. 2007):

$$\dot{Q}_{rad} = \sigma\varepsilon_r A_e F_r \left(T_{fg}^4 - T_s^4\right) \tag{3.18}$$

with ε_r the surface emissivity, σ the Stefan Boltzmann constant and F_r the radiation view factor. The radiation heat transfer rate is related to the radiation heat transfer coefficient by:

$$\dot{Q}_{rad} = u_{rad} A_e \left(T_{fg} - T_s\right) \tag{3.19}$$

Substituting equation (3.18) into equation (3.19) allows us to write the radiation heat transfer coefficient as:

$$u_{rad} = \varepsilon_r \sigma F_r \frac{T_{fg}^4 - T_s^4}{T_{fg} - T_s} \tag{3.20}$$

The high thermal conductivity of steel, along with the internal convection thermal resistance being lower than that of the external convection and radiation, allows the external surface temperature, T_s, to be well approximated by the steam temperature with clean tubes. These temperatures are therefore assumed to be equal for simplicity, although a more accurate value can be obtained iteratively once the conduction thermal resistance is known.

The radiation thermal resistance is calculated as follows:

$$R_{rad} = \frac{1}{u_{rad} A_e} \tag{3.21}$$

The resistance to heat transfer caused by the tube external ash layer, the steel wall and the internal oxide layer combines to give the total conduction thermal resistance as follows:

$$R_{cond} = \frac{\ln(r_1/r_i)}{2\pi k_{oxide} L} + \frac{\ln(r_2/r_1)}{2\pi k_{steel} L} + \frac{\ln(r_e/r_2)}{2\pi k_{ash} L} \tag{3.22}$$

with *k* the conductivity of each layer.

The fluid variables used as inputs to the above equations are determined by principles of mass, energy and momentum conservation for one-dimensional control volumes.

3.1.3 Conservation equations

The laws of mass, energy and momentum conservation govern the thermofluid interactions occurring within boilers. For one-dimensional steady state flow at constant elevation with no work being done, these laws can be formulated into the following equations

$$\dot{m}_{in} = \dot{m}_{ex} \tag{3.23}$$

$$\dot{m} h_{ex} - \dot{m} h_{in} = \dot{Q} \tag{3.24}$$

$$p_{ex} - p_{in} = \Delta p_L \tag{3.25}$$

The subscripts *in* and *ex* refers to the inlet and exit of the control volume. The difference in pressure caused by losses Δp_L must be characterised to solve momentum conservation as described by Equation (3.25).

3.1.4 Pressure losses

Fluid flowing in pipes or ducts is exposed to pressure losses. These losses are caused by friction between the fluid and the pipe surface (primary losses), turbulence induced by geometrical features in the flow path (secondary losses) and elevation differences (Rennels & Hudson 2012). The primary and secondary losses are functions of the dynamic pressure p_D of the fluid .

$$p_D = \tfrac{1}{2}\rho w^2 \tag{3.26}$$

A component specific pressure loss coefficient *K*, being the sum of primary and secondary loss coefficients, relates the dynamic pressure to the total pressure drop Δp through that component.

$$\Delta p = \left(K_P + K_S\right) p_D \tag{3.27}$$

Secondary loss coefficients are usually empirically developed for specific flow geometries and are available for commonly encountered components. Weisbach (1845) developed an equation for calculating the primary loss coefficient based on the size of the component and the Darcy-Weisbach friction factor *f*.

$$K_P = \frac{fL}{D_H} \tag{3.28}$$

An implicit empirical correlation known as the Colebrook-White formula can be used to numerically determine Darcy-Weisbach friction factor based on the surface roughness ε_s and the fluid flow characteristics (Colebrook 1939).

$$\frac{1}{\sqrt{f}} = -2\log\left(\frac{\varepsilon_s}{3.7D_H} + \frac{2.51}{\mathrm{Re}\sqrt{f}}\right) \tag{3.29}$$

3.2 Principal component analysis

The idea of finding the spaces of best fit to a dataset irrespective of the dependency amongst the variables was first presented by Pearson (1901). This concept was later named principal component analysis (Hotelling 1933) and has since found applications in data compression (Kambhatla & Leen 1997) and process monitoring and control (He et al. 2006).

PCA is a statistical multivariate data analysis technique which identifies orthogonal dimensions of descending variation in a dataset. These dimensions, defined by the correlation between variables, are called the principal components and are orthogonal and equal in number to the dimensionality of the dataset. The principal components can then be split into two subspaces called the principal component subspace (PCS) and the residual subspace (RS). The PCS is composed of an arbitrary

number of dimensions with the greatest variation and the RS of those with the remaining and least variation. By disregarding the RS, the data dimensionality can be reduced while retaining the most variation in the dataset. PCA is also useful for encapsulating the normal operating region using thresholds on the principal components instead of the original variables, providing a convenient way of detecting changes in the relationship between process variables.

A description of the PCA algorithm is now given. Let **X** be a $m \times n$ matrix, being the normalised values of a matrix with m samples of n variables. The principal component loading vectors of **X** are the eigenvectors of its covariance matrix $\mathbf{\Sigma}$.

$$\mathbf{\Sigma} = \frac{\mathbf{X}^T\mathbf{X}}{m-1} \tag{3.30}$$

The eigenvectors of $\mathbf{\Sigma}$ are its left-singular vectors being the columns of the matrix **U** obtained by the singular value decomposition (Robinson 2011) of $\mathbf{\Sigma}$.

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{3.31}$$

In equation (3.31), the columns of $\mathbf{V}^T$ are the right-singular vectors of $\mathbf{\Sigma}$. The diagonal elements of $\mathbf{\Lambda}$ are the ordered eigenvalues of $\mathbf{\Sigma}$. These eigenvalues represent the amount of variation along each corresponding eigenvector. These eigenvectors, being the principal component loading vectors, are an orthogonal basis set. The linear transformation of **X** onto this basis set is given by

$$\mathbf{T} = \mathbf{X}\mathbf{U} \tag{3.32}$$

where the matrix **T** contains the scores of each sample on the principal component axes.

By discarding the dimensions with the least amount of variation, the data can be compressed onto fewer dimensions while retaining the most significant variation in the original dataset. If the first *p* principal components are retained, the transformation of **X** onto the PCS and the RS is obtained using $\hat{\mathbf{U}} \in \mathbb{R}^{m \times p}$ and $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times (n-p)}$ respectively, with $\mathbf{U} = \begin{bmatrix} \hat{\mathbf{U}} & \tilde{\mathbf{U}} \end{bmatrix}$.

$$\hat{\mathbf{T}} = \mathbf{X}\hat{\mathbf{U}} \tag{3.33}$$

$$\tilde{\mathbf{T}} = \mathbf{X}\tilde{\mathbf{U}} \tag{3.34}$$

The PCS and the RS can now be projected back onto the original variables, decomposing the matrix **X** into its principal and residual components.

$$\hat{\mathbf{X}} = \hat{\mathbf{T}}\hat{\mathbf{U}}^T \tag{3.35}$$

$$\tilde{\mathbf{X}} = \tilde{\mathbf{T}}\tilde{\mathbf{U}}^T \tag{3.36}$$

with $\mathbf{X} = \hat{\mathbf{X}} + \tilde{\mathbf{X}}$.

3.2.1 Fault detection with PCA

The normal operating region is first identified using data representing a range of fault free operation. The principal components in which significant variation is expected to occur are then chosen as the PCS, resulting in a linear model for normal operation. If a new measurement sample **x** contains a fault that breaks the normal correlation between the measured process variables, it will cause a deviation from the PCS. Such faults will be observed as an increasing magnitude of $\tilde{\mathbf{x}}$, the residual between the actual measurement and its projection onto the PCS. One monitoring index for this deviation in the residual space is given by the squared prediction error, *SPE*.

$$SPE = \tilde{\mathbf{x}}^T\tilde{\mathbf{x}} \tag{3.37}$$

Some faults do not necessarily disturb the normal correlation between process variables but are caused by process boundary conditions exceeding the normal threshold. These faults are manifested as large scores within the PCS and can be monitored using Hotelling's T-squared statistic, *TS*.

$$TS = \mathbf{x}^T\hat{\mathbf{U}}\Lambda^{-1}\hat{\mathbf{U}}^T\mathbf{x} \tag{3.38}$$

A fault is detected when these indices exceed a specified threshold. Yue & Qin (2001) suggest the use of a combined index, φ, which unifies deviations in the PCS and RS.

$$\varphi = \frac{SPE}{\delta_\alpha^2} + \frac{T^2}{\chi_\alpha^2} \tag{3.39}$$

In equation (3.39), δ and χ are the thresholds for *SPE* (Nomikos & MacGregor 1995) and *TS* (Qin 2003) respectively, based on a specified confidence interval α. The combined index essentially encapsulates the normal operating region (NOR).

3.2.2 Fault isolation with PCA

Contribution plots

Fault isolation with contribution plots is based on establishing the contribution of each measured variable to the deviation from the PCS. The contribution of variable *v* is given by the square of the corresponding entry in $\tilde{\mathbf{x}}$ (Miller et al. 1998).

$$Cont_v = \left(\tilde{\mathbf{x}}_v\right)^2 \tag{3.40}$$

Reconstruction-based contribution

The concept of reconstruction proposed by R Dunia & Qin (1998) was later combined with contribution plots (Li et al. 2014) to form the reconstruction-based contribution (RBC) method

proposed by Alcala & Qin (2009). The contribution of a variable to the RBC index is the amount of reconstruction along a fault direction required to reconstruct the faulty sample back into the NOR. The advantage of RBC over reconstruction is that it is not limited to single variable faults (Alcala & Qin 2009) and it gives improved diagnosis results (Li et al. 2014).

The RBC index for faults occurring in the RS is obtained using a fault library $\boldsymbol{\Xi}$ with *i* columns, each column being a different fault vector.

$$RBC_i = \mathbf{x}^T \tilde{\mathbf{C}} \boldsymbol{\Xi}_i \left(\boldsymbol{\Xi}_i^T \tilde{\mathbf{C}} \boldsymbol{\Xi}_i \right)^{-1} \boldsymbol{\Xi}_i^T \tilde{\mathbf{C}} \mathbf{x} \tag{3.41}$$

with $\tilde{\mathbf{C}} = \tilde{\mathbf{U}} \tilde{\mathbf{U}}^T$.

For unidimensional faults, that is, faults only affecting a single measured variable, equation (3.41) reduces to

$$RBC_i = \frac{\tilde{\mathbf{x}}_i^{\,2}}{\tilde{\mathbf{C}}_{i,i}} \tag{3.42}$$

where $\tilde{\mathbf{C}}_{i,i} = \boldsymbol{\Xi}_i^T \tilde{\mathbf{C}} \boldsymbol{\Xi}_i$ is the i^{th} diagonal element of $\tilde{\mathbf{C}}$.

4. Methodology

The problem of online monitoring is shaped by the system, its measurements, and monitoring objectives. The online monitoring frameworks discussed in this section are therefore developed with specific reference to the layout and measurement configuration of a case study boiler. A description of the case study boiler is given, followed by the development of online monitoring frameworks for the convective pass heat exchangers respectively based on soft sensor design and data-driven monitoring. The soft sensor approach deals with the design of a model that estimates the state of unmeasured variables related to the health and performance of the boiler heat exchangers. The data-driven approach identifies patterns in measurement data indicating a deviation of specific variables from normal. Both methods only make use of the standard process measurements installed on the case study boiler. The benefits, limitations and achievable monitoring outcomes of each approach are compared.

4.1 Case study boiler description

The online monitoring frameworks are based on the flow and measurement configuration of an existing 620 MW coal fired two-pass drum boiler. Its convective pass consists of one economiser stage, three superheat stages (SH1, SH2 and SH3) and two reheat stages (RH1 and RH2). The arrangement of these heat exchanger stages is shown in Figure 4-1.

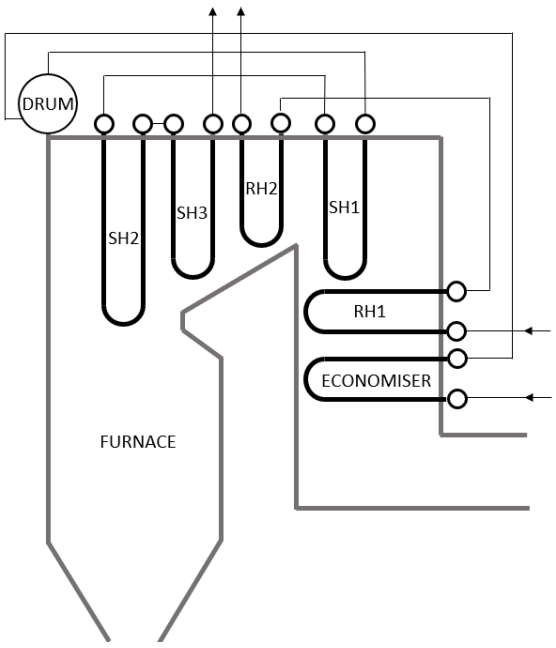


Figure 4-1: Side view of the heat exchanger stage arrangement in the convective pass of the case study boiler.

The steam flow in the superheat and reheat stages are split into four legs (A, B, C, D) running in parallel across the boiler width. As the steam flows from one stage to the next, these legs alternate the steam flow between the sides and across the centre of the boiler to disperse the effects of an uneven heat profile along the cross section of the flue gas flow path as represented in Figure 4-2. This is the custom practice in modern boiler designs (Nielson, F.S., Danesi, P. and Radhakrishnan 2012). Although each of these legs is composed of several tube elements, each one will be represented as a single lumped heat exchanger throughout this study.

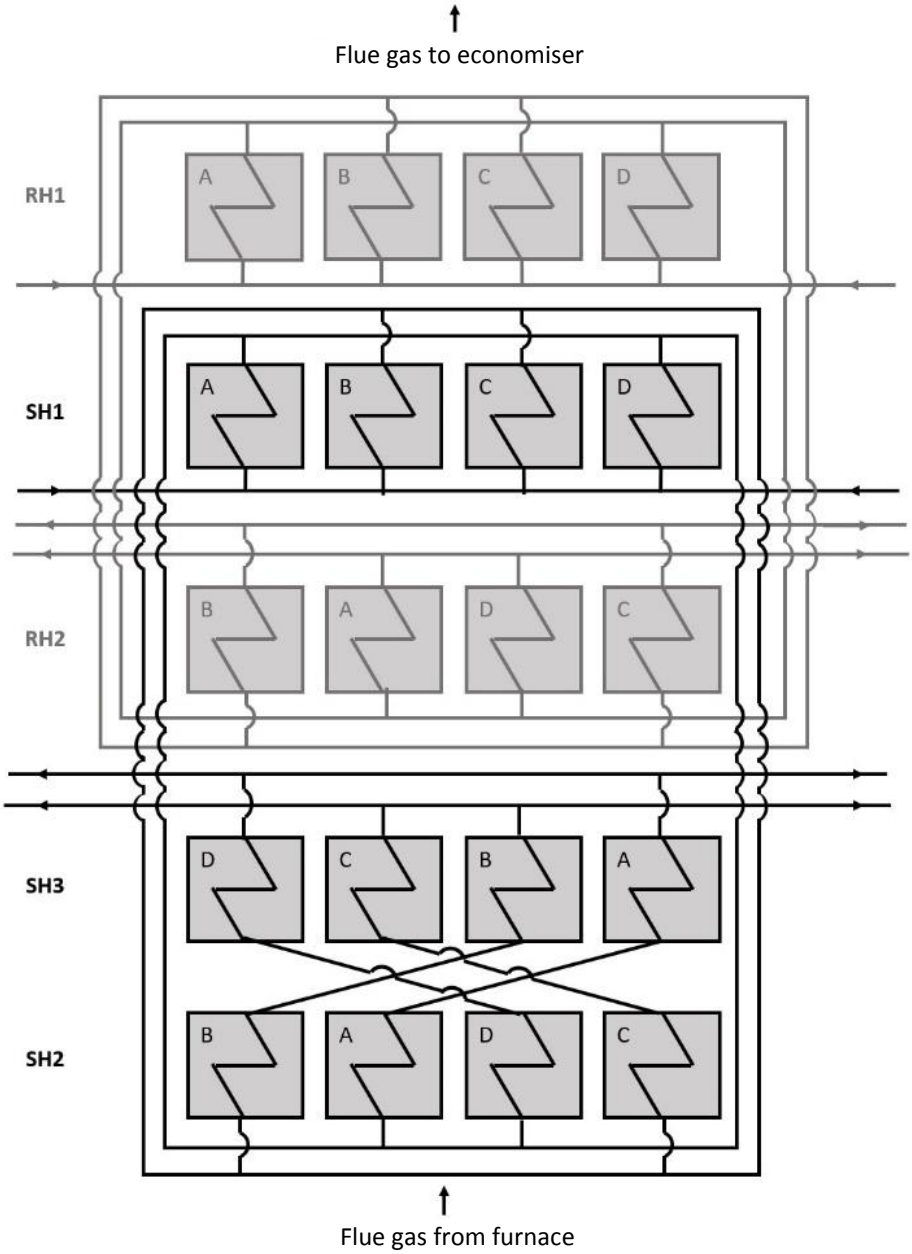


Figure 4-2: Case study boiler convective pass heat exchanger leg layout.

Spray water attemperators are installed on each of the four interconnecting pipes before the second and third superheat stages and before the second reheat stage, creating three attemperator stages consisting of four attemperators each. Two-stage attemperation is used for the superheat section, employing coupled two-stage control, in which the first stage attemperators maintain a specified temperature difference across the corresponding second stage attemperators, which in turn responds mainly to the final outlet steam temperature measurements on the respective legs. The reheat steam outlet temperature is controlled by single-stage attemperation.

The account of the boiler measurements to follow is not exhaustive, limited to those deemed relevant. The steam measurement configuration for each heat exchanger stage is shown Figure 4-3. Note that legs A and D are connected by common headers throughout the respective superheat and reheat cycles, and similarly legs B and C. Only the combinations of legs A and D are included in Figure 4-3, as the same holds for the combinations of legs B and C. Steam temperature measurements are located at the inlet to both the superheat and reheat sections, upstream and downstream of each attemperator, and at the final outlet of both the superheat and reheat sections on the individual legs. Steam flow measurements are located at the final superheater outlet only, measured on each individual leg. Spray water mass flow rates through each attemperator is measured. The temperature of both the superheat and reheat spray water is also measured for the respective tap off points from the feed water system.

Flue gas temperature measurements are located both upstream and downstream of the economiser. Flow rates of coal and air into the boiler are also measured.

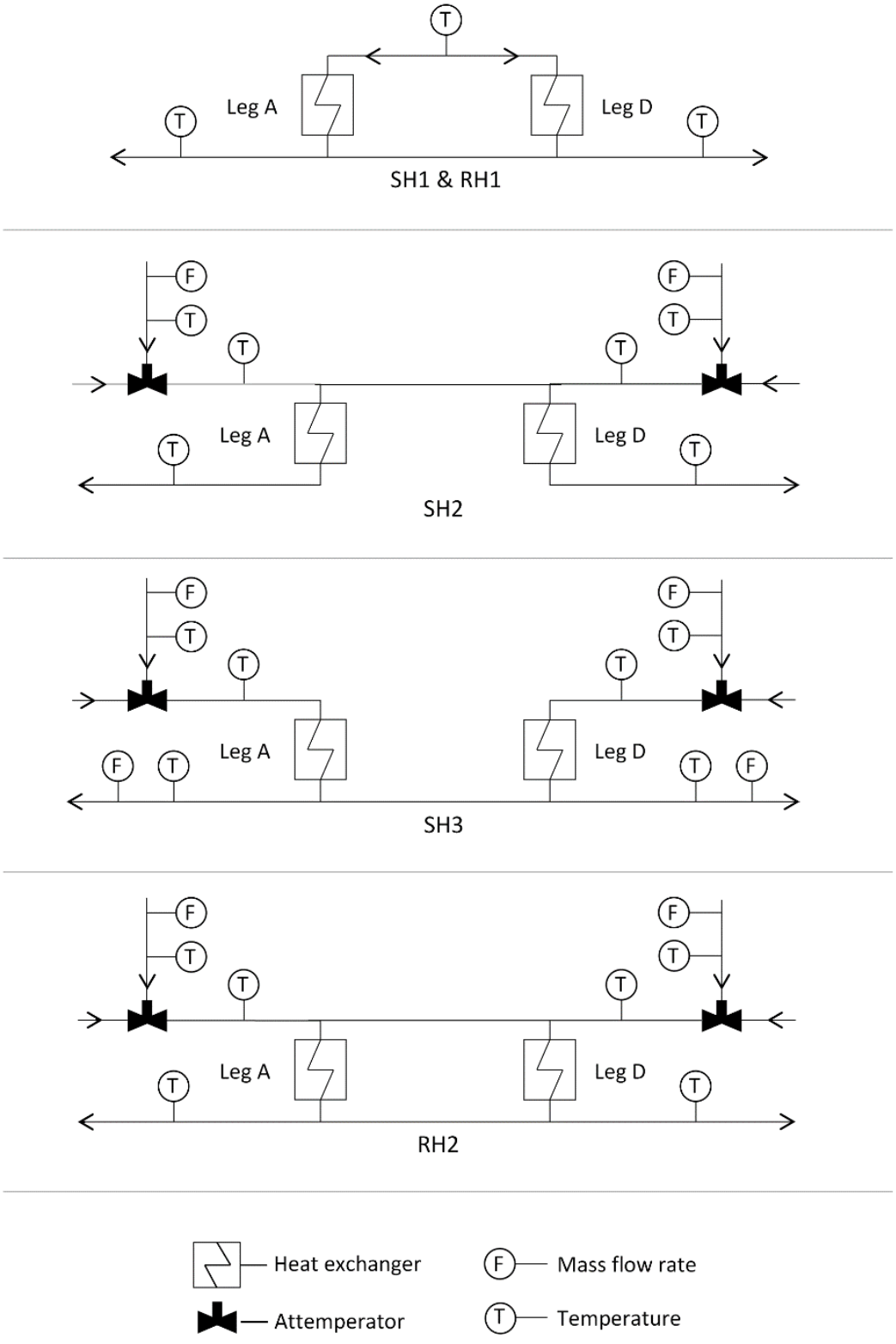


Figure 4-3: Steam side measurement configuration for the respective heat exchangers

4.2 Soft sensor design

Models devoted to the estimation of plant variables are known either as inferential models, virtual sensors or soft sensors (Fortuna et al. 2007) and will be referred to as soft sensors throughout this study.

The soft sensors found in literature for boiler heat exchanger monitoring are designed using mass and energy conservation models to achieve online monitoring of heat exchanger stages. The flow and measurement configuration of the case study boiler shown in Figure 4-2 and Figure 4-3 potentially lends itself to increasing the monitoring resolution to each heat exchanger leg. However, the steam flow rates through each leg, and the inlet and outlet steam temperatures of each leg, are not immediately evident from the measurements because mixing can occur in the common headers connecting pairs of heat exchanger legs. Mixing can also occur in the flue gas flow path, and the flow distribution is required to obtain the average flue gas temperatures surrounding each leg. Alternatively, a known temperature distribution enables the average flue gas flow through each leg to be estimated. These parameters are required to monitor individual legs using mass and energy conservation models. By additionally solving for momentum conservation while taking into account the geometry and pressure losses through components, the actual flow rate in connecting headers and ducts can be estimated, allowing each leg to be monitored individually. This will enhance the monitoring resolution and therefore the extent to which the outcomes of monitoring can be optimized.

The soft sensor design for this case study boiler is based on a one-dimensional, steady state, thermofluid model. The model is setup to take actual plant measurements as inputs from which it approximates the thermodynamic state of both the steam and flue gas throughout the convective pass in accordance with the model's level of discretization.

4.2.1 Steam side model

The steam side is modelled according to the flow layout shown in Figure 4-2 with steam as the working fluid. Each heat exchanger leg is modelled as a node that receives heat from the flue gas. The dimensions and other physical characteristics of the surrounding pipes and flow connections are taken from plant design data to account for flow resistance as a function of flow characteristics. In addition to the available measurements shown in Figure 4-3, the total reheat steam mass flow rate is required to fully determine the model. This quantity is not measured directly and can be soft sensed by a mass balance around the high pressure turbine as described by ASME (2008).

Using the estimated total reheat steam mass flow rate along with the hard measurements as input boundary conditions, the model determines the required rate of heat transfer into each leg to match the boundary conditions while conserving mass, energy and momentum. The results therefore also include the thermodynamic state of the steam at each location throughout the final superheater system with accompanying flow distributions.

4.2.2 Flue gas side model

The flue gas side is modelled according to the flow layout shown in Figure 4-4. The flue gas properties of enthalpy and density are calculated according to the description in Section 3.1.1 using the typical flue gas composition of the case study boiler at full load. Although the flue gas composition is a function of load, the difference in these property relations at different loads is minor and a constant flue gas composition is therefore assumed. Curves showing these properties for a range of temperatures, pressures and load conditions are included in Appendix A with a comparison to air. All other flue gas properties are assumed to be those of air. Unlike the steam side model flow layout which is based on the flow paths created by the physical pipe layout, the flue gas model layout splits the flow into four virtual flow paths corresponding to the four heat exchanger legs of each stage. This level of model discretization allows for evaluating the average flue gas properties for each leg.

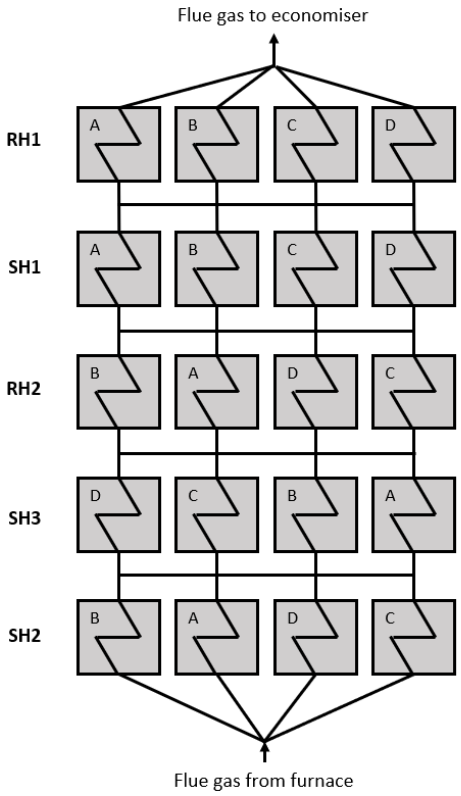


Figure 4-4: Flue gas model arrangement

As with the steam side, each heat exchanger leg is modelled as a node from which heat is extracted at a rate corresponding to the calculated rate of heat input to the steam. The effects of flue gas heat losses and directed radiation are therefore neglected. The characteristics of the pipe and remaining flow connections are chosen to represent the flue gas flow path resistance. The flows are completely connected between each heat exchanger stage with pipes dimensioned to represent very low flow resistance as the flow areas here are essentially large cavities with no physical boundary separating the four flow streams. These connections allow mixing to take place according to property differences caused by uneven heat extraction or flow maldistribution through the four different legs.

In addition to the available measurements detailed in Section 4.1, the total flue gas mass flow rate is required to fully determine the model. This quantity is not measured directly and can be soft sensed using a mass balance around the furnace (ASME 2008), or an energy balance around the economiser (Shi et al. 2015). The latter approach is employed in this study and the reason for this choice is elaborated in Appendix B.

Using the estimated total flue gas mass flow rate, the temperature measurement at the economiser inlet and the rate of heat transfer to each leg as input boundary conditions, the model solves for mass, energy and momentum conservation. The model consequently computes the thermodynamic states throughout the flue gas side along with the accompanying flow distributions according to the model layout. The appearance of the combined steam and flue gas side models in the actual modelling software is shown in Appendix C.

4.2.3 Heat transfer resistance analysis

The soft sensor developed up to here provides, for each heat exchanger leg, the average steam and flue gas properties and flows, along with the heat transfer rates. With these parameters available, the soft sensor is now extended to evaluate the different components of the overall thermal resistance of a specific heat exchanger leg.

First the overall thermal resistance of each leg is set to achieve the full load design heat transfer rates at the corresponding temperature differences and flow rates using equations (3.5) to (3.10). The overall thermal resistance is composed of the sum of the radiation, convection and conduction thermal resistances, which are in turn functions of the fluid variables and the geometry and composition of the heat exchanger. It is assumed that no fouling is present at the design conditions and that the overall thermal resistance is only due to flue gas radiation and convection. If a baseline fouling was however incorporated in the design, this approach will produce the relative changes from this baseline. By now specifying the contribution of radiation to

the total heat transfer rate of a heat exchanger, the design convection heat transfer resistance can be obtained. Furthermore, the value of the $\sigma\varepsilon_r A_e F_r$ term in equation (3.18) can be determined, which is assumed to remain constant.

During operation, the actual overall thermal resistance is calculated by again using equations (3.5) to (3.10). The convection thermal resistance for off design operation is then determined by scaling the design convection resistance by the deviation of the fluid variables from their design condition according to their orders in equations (3.14) to (3.17):

$$R_{conv} = R_{conv_D} \frac{k_D \left(\rho_D w_D \mu_D{}^{-1}\right)^m \left(c_D \mu_D k_D{}^{-1}\right)^{0.33}}{k\left(\rho w \mu^{-1}\right)^m \left(c \mu k^{-1}\right)^{0.33}} \tag{4.1}$$

The subscript D refers to the design value of the variable. Equation (4.1) can be reduced to:

$$R_{conv} = R_{conv_D} \left(\frac{\dot{m}_D}{\dot{m}}\right)^m \left(\frac{k_D}{k}\right)^{2/3} \left(\frac{c_D}{c}\right)^{1/3} \left(\frac{\mu_D}{\mu}\right)^{1/3-m} \tag{4.2}$$

Rather than updating the value of *m* in equation (4.2) with the Reynolds number according to Table 3-1, a constant value of 0.6 is used, which has been shown to be valid for CFPP convective heat exchangers (Grimson 1937).

By updating equation (3.20) with the actual temperatures, the radiation resistance can also be calculated for off design conditions.

The approach above allows the convection and radiation resistance to be estimated as a function of the fluid variables without detailed information about the heat exchanger geometry. The implementation of this approach in the modelling software is verified with a hand calculation included in Appendix D.

Using the actual fluid variables, the real time convection and radiation components of heat transfer resistance for each heat exchanger leg is therefore calculated by assuming a constant heat exchanger geometry and composition. The same assumption will not hold when evaluating the conduction thermal resistance as it is mainly influenced by the thickness of the ash and oxide layers on the tubes. However, the difference between the overall thermal resistance and those portions of it not including the conduction thermal resistance yields the conduction thermal resistance:

$$R_{cond} = R_{overall} - \left(\frac{1}{R_{conv_e}} + \frac{1}{R_{rad}}\right)^{-1} \tag{4.3}$$

By directly monitoring the conduction resistance, the effects of ash and oxide layer build-up on the heat transfer rates are disconnected from those of changes in the operating conditions.

4.2.4 Model validation

The steam and flue gas models are now simulated with inputs corresponding to full load design conditions. The model outputs of interstage flue gas temperature are compared to the expected values given in the design information as shown in Figure 4-5. The line of exact correspondence is indicated by y=x.

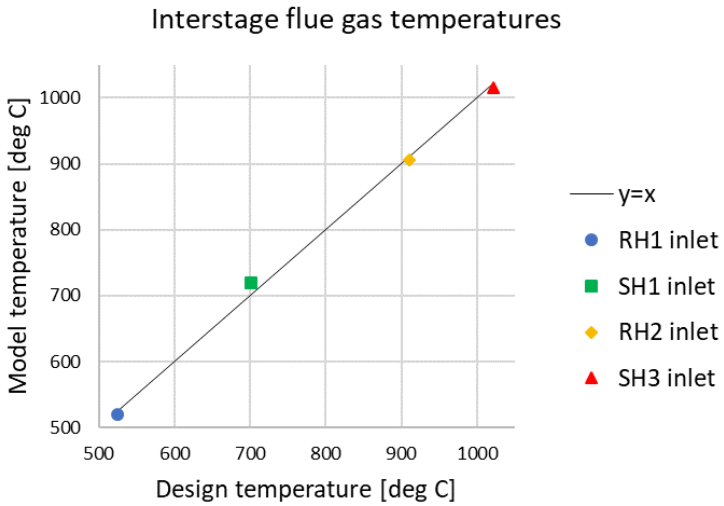


Figure 4-5: Interstage flue gas temperatures comparison to design values

A difference is observed between the model outputs and design information. This confirms that the developed model differs from the model used to design the boiler. This is expected because of the simplifying assumptions made in the model development. The model outputs are however in reasonable agreement with the design information, indicating that the developed model is similar to the design model.

The overall thermal resistance of each heat exchanger leg is set to achieve the full load design conditions assuming no conduction resistance. It follows that the conduction resistance calculated for these conditions should therefore be close to zero. If the design interstage flue gas temperatures are used to set the overall resistance, the conduction resistance is however expected to deviate from zero because of the discrepancy in model outputs shown in Figure 4-5. The model outputs for the interstage flue gas temperature are therefore used instead of the design values to set the overall resistance of the heat exchangers.

Full load corresponds to a boiler maximum continuous rating (BMCR) of 97%. The design information also contains information for 70% and 100% BMCR and the model is now also simulated for these design inputs. Although the actual overall resistance of each heat exchanger calculated at the different loads will change from the value set for full load, it is expected that the convection and radiation resistance should scale in accordance with the changing process conditions, resulting in the conduction resistance remaining close to zero. This is assuming that the same conduction resistance was used for different loads in the design. This will test if the convection and radiation resistance correlations appropriately respond to the changing process conditions. The overall resistance, along with the expected convection and radiation resistance for leg A only are plotted in Figure 4-6 for the three different loads.

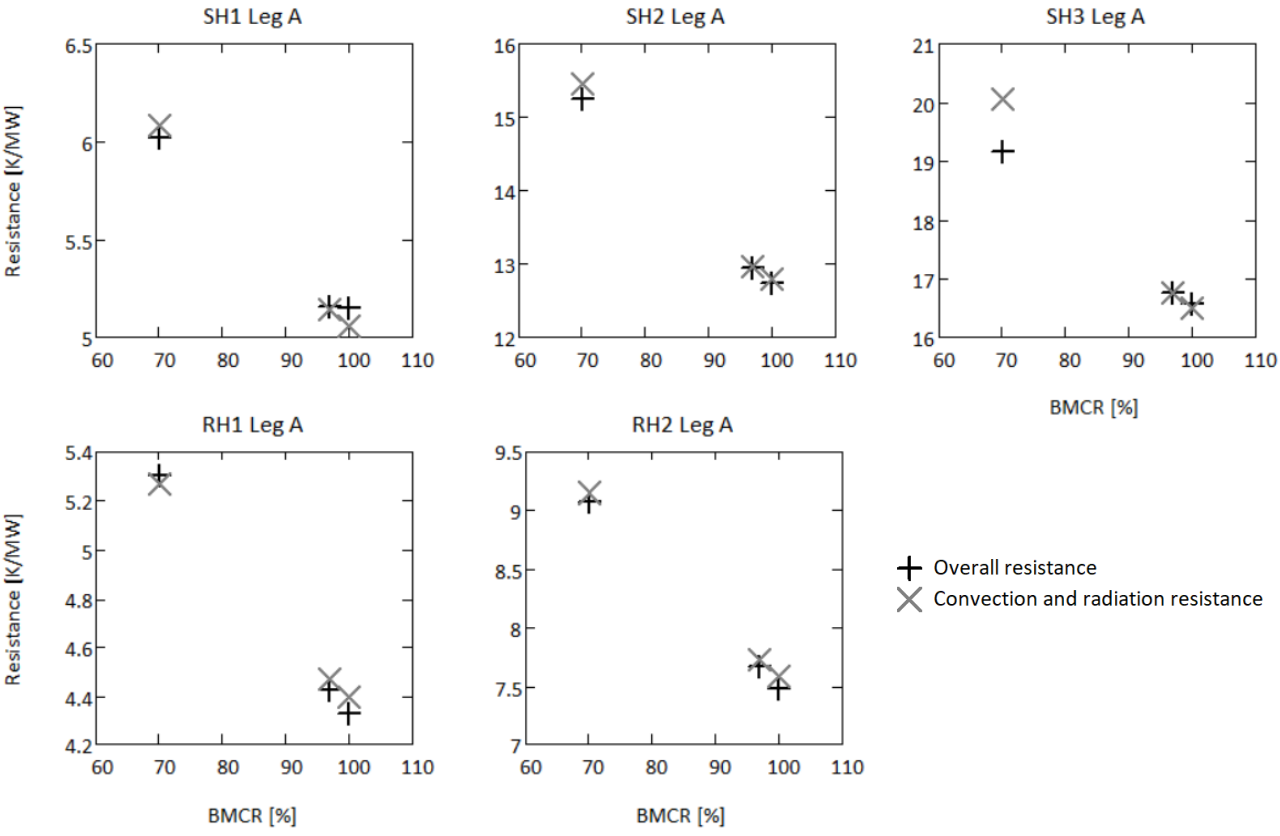


Figure 4-6: Scaling of convection and radiation resistance at different loads to match overall resistance

The conduction resistance is the difference between the overall resistance and the convection and radiation resistance as per equation (4.3). The convection and radiation resistance components do not correlate perfectly with the overall resistance at the different loads. This is probably because the correlations differ from that used for the design of the boiler. The contribution of conduction

resistance however remains small in relation to the overall resistance. The model is therefore considered valid for the objectives of this study and no further calibration is done.

4.3 Data-driven monitoring

4.3.1 Background

Data-driven monitoring aims to identify patterns in measurement data indicative of the process state (Yin et al. 2012). It is an attractive alternative to model-based monitoring when process measurements are not enough to fully determine the achievable process models, or when the model complexity required to do so is too great to bear computationally in online applications.

Literature pertaining to data-driven monitoring of boilers are focussed on preventing boiler trips by proactively alerting operators about conditions alike to that experienced at the onset of previous trips (Yu, Jang, Yoo, Park, et al. 2016; Ismail et al. 2016; Yu, Jang, Yoo & Kim 2016; Yu et al. 2015). This prompts the operator to investigate the cause and take preventative action. The investigative task is challenging as the actual cause is often obscured by the interconnectedness of boiler measurement response to faults. It is therefore desirable to deepen the diagnosis to monitor the condition of individual components to not only prevent trips, but to identify conditions detrimental to the reliability and performance of boiler components.

Selection of a data-driven monitoring method is highly dependent on the type of system, its measurement configuration and the type of faults to be identified. Boiler damage is driven by several mechanisms being simultaneously present in different proportions according to the influence of operating actions on process conditions. For practical boiler monitoring, a method capable of ascribing the interconnected response of measurements to different faults is required. Data-driven diagnostic classifiers like RBC and ANNs cannot identify multiple simultaneous faults (Venkatasubramanian, Rengaswamy, Kavuri, et al. 2003). These methods can only identify a combination of faults if that specific combination is included in the fault matrix or training data. This essentially means that each combination of faults to be diagnosed must be presented to the classification algorithm as an individual fault during its training. This might be feasible by discretizing the continuous spectrum of all fault combinations into a manageable amount, given that there exists a way to easily control fault variables and generate fault data. The number of additional faults created by this approach however increases exponentially with the number of faults. For a large system this can quickly become “combinatorially prohibitive” as stated by Venkatasubramanian, Rengaswamy, Kavuri, et al. (2003).

A data-driven method is therefore developed to suit thermofluid process behaviour like that encountered in boiler heat exchangers. Before describing the method and its application to the case-study boiler, a simplified heat exchanger system model will first be simulated to describe some fundamentals of the data-driven monitoring problem and to share some relevant steps that led to the development of the data-driven boiler monitoring method.

4.3.2 Simplified heat exchanger system analysis

The simplified heat exchanger system consists of two convective heat exchangers (HE1 & HE2), connected as shown in Figure 4-7.

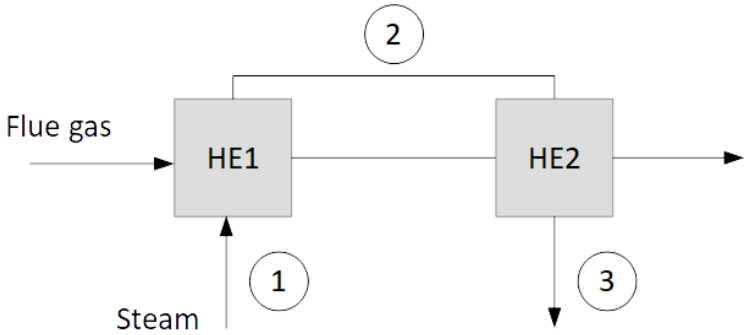


Figure 4-7: Simplified heat exchanger system layout

The system is modelled to account for heat transfer due to tube external convection and radiation with the applicable correlations described in Section 3.1 and conduction through the tube and fouling material. It is assumed that the only measurements taken on this system are the steam temperatures at points 1, 2 and 3.

First the model is simulated with constant process inputs and with the conduction thermal resistance of both heat exchangers corresponding to normal operation. Faults are then imposed by introducing an incremental increase in the conduction resistance of the respective heat exchangers. In Figure 4-8, the two-dimensional projections of the three-dimensional steam temperature dataset is plotted onto the $T_1 - T_2$, $T_2 - T_3$ and $T_1 - T_3$ planes. The faults are evident in the measurement data by its deviation from the normal operating point (NOP). The data corresponding to the greatest deviation in conduction thermal resistance of the heat exchangers are the furthest away from the NOP, plotted in bold. For the current process inputs and the magnitude of the imposed faults, the direction of movement in the measurement space appears to be linear both in the direction and rate of propagation.

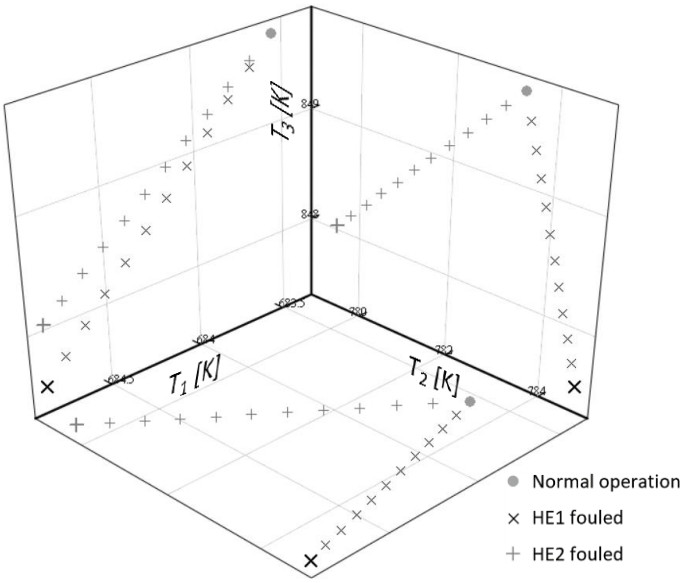


Figure 4-8: Two-dimensional steam temperature plots at constant input conditions

The model is now simulated for different operating conditions to imitate data obtained from actual plant operation. The load is varied by simultaneously adjusting all process inputs to match the design conditions corresponding to the load. Process noise is added by allowing all inputs to randomly deviate from the design condition at the corresponding load within plausible limits. Measurement noise is added to the output data to reflect inaccuracies in temperature measurement instrumentation. The simulation is run for normal operation and then repeated for a step increase in the conduction thermal resistance of the respective heat exchangers. The resulting data is plotted in Figure 4-9.

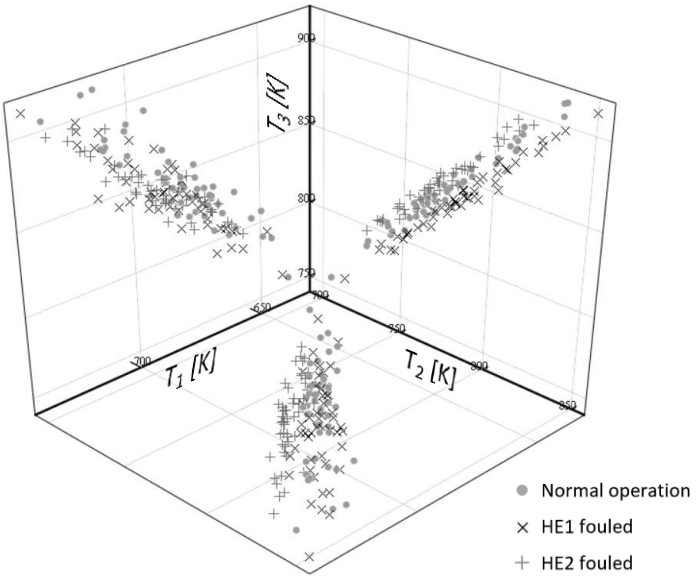


Figure 4-9: Steam temperature plots for a range of loads with the addition of process and measurement noise

No clear distinction between the normal operation and degraded operation data can be observed when plotted against any pairs of steam temperatures. The separation of the faults from the normal operating regions in the measurement data is obscured by the addition of noise and load variations.

Principal components analysis (PCA) as described in Section 3.2 is now conducted on the dataset corresponding to normal operation. The data obtained from the previous simulation is transformed onto the principal components and plotted in Figure 4-10.

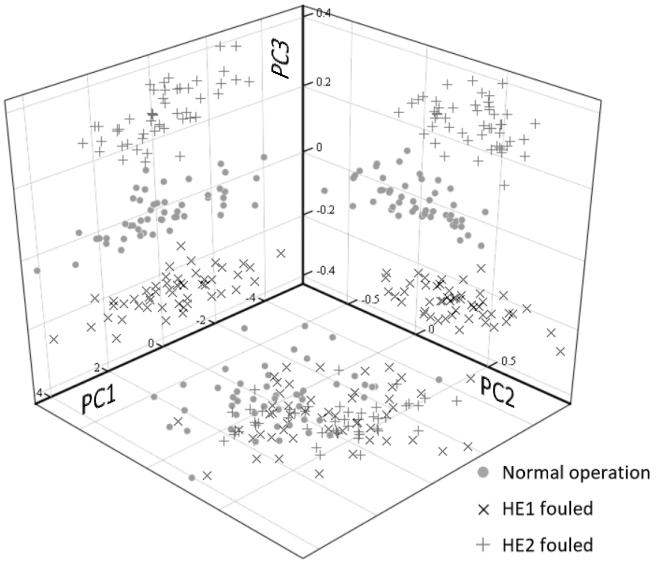


Figure 4-10: Steam temperature data plotted on the principal components

While the first principal component contains almost no information that distinguishes the normal from the faulty data, and the second principal component contains very little, a clear distinction is seen in the third principal component. It is expected with thermofluid processes that faults are manifested in the last few principal components, as they cause deviations from the correlation between process measurements at normal operating conditions. It must however be noted that the separation between the normal and faulty data observed in the principal component space also exists in the normal measured parameter space. This separation is simply visualized better in two dimensions in the principal component space. The statistical confidence with which one can say that the data falls outside of the normal operating region depends on a combination of the magnitude of the fault and the ranges of process and measurement noise.

In boilers, it can seldom be expected that a single heat exchanger will deteriorate in isolation while others remain healthy. The above simulation is therefore repeated to observe the effects of the two faults acting simultaneously in Figure 4-11.

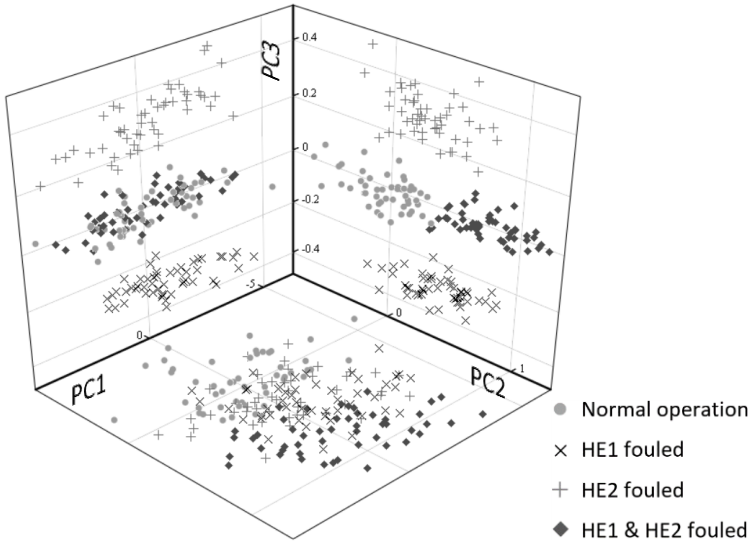


Figure 4-11: Multiple simultaneous faults plotted in the principal component space

The effects of the two individual faults are superimposed when they are simultaneously present. The combination of these faults is most evident in the second principal component.

For practical boiler monitoring, a method is required that is applicable to a range of loads, while still being able to uniquely diagnose multiple simultaneous faults. The description of such a method is given below.

4.3.3 Data-driven method

For thermofluid processes represented by continuous, smooth and orderly functions, continuous incremental changes in unmeasured variables, which have an influence on measurements, will cause continuous incremental movement in the measurement space. The direction of this movement is defined by the influence of the changing variable on measured variables. Assuming this movement is linear, and its direction is not affected by the state of other variables, this direction serves as a signature corresponding to a change of that specific unmeasured variable. The distance covered in this direction is related to the magnitude of the deviation. When multiple variables deviate simultaneously, their influence on the measurement space is superimposed.

A deviation from design in any parameter of monitoring interest can be considered as a fault (Isermann 2011). The direction vector corresponding to the measurement space movement caused by a fault is the fault signature. The fault vector can be obtained by subtracting the NOP from the measurements taken in the presence of the fault.

$$\mathbf{f} = \mathbf{x}_{fault} - \mathbf{x}_{normal} \tag{4.4}$$

The fault matrix is the collection of fault vectors corresponding to different faults.

$$\mathbf{F} = \left[\mathbf{f}_1, \ldots, \mathbf{f}_n\right] \tag{4.5}$$

During monitoring, the actual deviation vector is obtained by centering the measured sample around the NOP:

$$\mathbf{f}_{actual} = \mathbf{x}_{measured} - \mathbf{x}_{normal} \tag{4.6}$$

By changing the basis of the actual deviation vector using the fault matrix as the transformation matrix, the relative contribution of each fault to the measurement sample can be captured in a fault index.

$$\mathbf{i} = \mathbf{F}^{-1}\mathbf{f}_{actual} \tag{4.7}$$

This obviates the need for linearly independent fault vectors which spans the measurement space. In cases where the fault matrix is rank deficient, the pseudoinverse (Rao 1967) gives the best approximate solution to equation (4.7).

$$\mathbf{F}^{+} = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T \tag{4.8}$$

with $\mathbf{F}^{+}$ the pseudoinverse.

If the normal value of a variable corresponding to fault *j* is $v_{j_{normal}}$ and the value to which it was set during training is $v_{j_{fault}}$, the actual value of this variable during operation is given by:

$$v_{j\,actual} = v_{j\,normal} + \mathbf{i}_j\left(v_{j\,fault} - v_{j\,normal}\right) \tag{4.9}$$

4.3.4 Fault matrix development

A fault signature is an indication of the presence of a fault. The fault must therefore cause a unique and detectable response in measurements to have a fault signature. Fault signatures can be exposed through the analysis of labelled training data. Labelled data refers to data which includes both measured variables and their accompanying fault status. The process of correlating the fault status to measurement data is called training. Labelled data can be obtained from either historical process measurements or model simulation. The data presented in Figure 4-8 to Figure 4-11 are examples of labelled data generated by model simulation.

Obtaining labelled data from an operating boiler poses several limitations:

1. To get data corresponding to a known online fault status implies the existence of a means to determine its fault status online. This is difficult for power plant boilers because the harsh environment in which they operate impedes measurements for verifying the fault status.
2. Imposing faults in an actual boiler for training purposes is not feasible as the relevant fault variables are neither measured nor controlled.
3. Boilers are prone to several faults being active simultaneously, making it difficult to knowingly gather data only associated with a single fault.
4. Service intervals are quite dispersed, reducing the opportunity for correlating data of the previous operating cycle to the fault status established from offline inspection, limiting the amount of available training data.
5. Training data is limited to encountered faults.

Labelled data can also be generated using a model. One advantage of using a model is that the fault status can be completely controlled for all operational states. The accuracy of the data depends directly on how well the model mimics the actual plant response to faults.

A thermofluid process model is used to generate labelled training data for constructing the fault matrix of the case study boiler.

a) Model development

The same model developed and validated for the soft sensor is used for training the data-driven method. However, the input-output configuration differs from that of the soft sensor model. The soft sensor model receives measurements as inputs to calculate unmeasured variables. For constructing the fault matrix, unmeasured variables are used as inputs to calculate the expected measurements. This allows one to impose changes to the unmeasured variables of which a deviation from normal is considered as a fault and observe the consequent measurement response. Some measured process boundary conditions are however included as inputs to fully determine the model. These include the steam flow rates through the superheat and reheat sections, the steam temperature into the superheat and reheat sections, and the flue gas mass flow rate.

Instead of calculating the conduction thermal resistance of each leg as with the soft sensor, it can now be specified. This is added to the convection and radiation resistance, calculated using the same methodology as the soft sensor model, to obtain the overall thermal resistance. The

attemperator spray water mass flow rates are calculated to achieve the steady state steam temperature control setpoints.

b) Training methodology

The model is simulated for normal operation at full load conditions. The outputs of expected measured variables represent the normal operating point (NOP). Faults are then simulated by respectively imposing a deviation in the model input variables of which a change represents a fault. The resultant output data corresponding to each fault is centred around the NOP. This produces a direction vector corresponding to a fault which serves as the fault signature, $\mathbf{f}$, in equation (4.4). The deviation does not need to be imposed both in the positive and negative direction, but only in one direction. This is because the fault vector corresponding to a decrease in a specific variable is the negative of the fault vector for an increase in that same variable. Only one simulation is therefore needed per fault case.

The direction of the fault vector is however prone to change if there are nonlinearities in the measurement response to varying fault magnitudes, process boundary condition changes and the influence of other faults. If the effect on the fault vector direction is significant, diagnosis results may be inaccurate, and these aspects must therefore be considered in the training methodology.

Before evaluating fault vector linearity, the types of faults to be considered must be chosen. The faults included in the fault matrix greatly impacts the diagnostic capabilities of the data-driven method as diagnosis is essentially provided by simultaneously considering all faults to disconnect their interdependency.

Fault and measurement selection

Consider the case study boiler operating at fixed process boundary conditions with no external flows entering the system and with a fixed geometry. In this case the measurements can only be altered by changing conduction thermal resistance by fouling or oxide scaling. The resolution to which the location of changing conduction resistance can be diagnosed is that for which the fault vector direction produced for each different fault location is linearly independent. This is because a unique solution to equation (4.7) or equation (4.8) can only be expected if the fault vectors in the fault matrix are linearly independent, meaning that the measurement space movement can only be described by one combination of fault vectors. It is intuitive that the fault direction for changing conduction resistance anywhere on a specific leg will produce the same fault vector direction, irrespective of where on the leg the change takes place. This is because the measurement setup is such that it can at most provide information about the lumped properties

of each leg. Using this as a starting point, the fault vectors for changing conduction resistance of individual heat exchanger legs are studied.

The analytical evaluation of linear dependency is achieved by analysis of the singular values of a matrix. The matrix is said to be linearly dependent or singular if not all singular values are greater than zero. Since even slight nonlinearities in the direction of measurement response to faults will result in all singular values of the fault matrix taking on values greater than absolute zero, a numerical analysis of linear dependency is appropriate. One parameter that can be used for such an analysis is called the condition number of a matrix, which is the ratio of the maximum to the minimum singular values (Trefethen & Bau III 1997). A high condition number indicates that a matrix is near singular or ill-conditioned. Another approach is to directly observe the minimum singular value, where very small values indicate a near singular matrix. There are however no clear thresholds at which these parameters confirm linear dependency.

A different approach for evaluating linear dependency is therefore proposed. By diagnosing the fault matrix with itself, it is expected that each entry should in turn be uniquely diagnosed as the corresponding fault if the fault matrix is even barely linearly independent. In other words, the index matrix resulting from such a diagnosis of a non-singular matrix is an identity matrix.

$$\mathbf{F}^{-1}\mathbf{F} = \mathbf{I} \tag{4.10}$$

By however diagnosing a fault matrix with randomly imposed disturbances in the direction of each fault vector with the original fault matrix, the resulting index matrix is expected to deviate from the identity matrix. For a near linear dependent fault matrix, a small disturbance angle can cause the index matrix to show no resemblance to the identity matrix. For a fault matrix that is far removed from linear dependency, a big disturbance angle can still produce an index matrix that resembles the identity matrix. The extent of linear dependence can therefore be evaluated based on this deviation at a corresponding angle of the fault vector disturbance. This method of linear dependence evaluation is referred to as a reciprocal analysis as per equation (4.10) with a specified disturbance angle throughout this study. The disturbance is imposed such that all fault vectors in the fault matrix randomly deviates in any direction at a specified angle.

The minimum number of measurements required to achieve linear dependency is evaluated. It is also known that neither equation (4.7) nor equation (4.8) will be solvable if the number of measurements are not greater than or equal to the number of faults. There are 20 individual conduction resistance faults, one for each leg. Therefore, at least 20 measurements must be considered. It is intuitively known that the measurements of inlet boundary values are not

influenced by changes in conduction thermal resistance. Therefore, only interstage steam temperature, spray water flow, and flue gas outlet temperature measurements are considered.

It is further observed that the measurements of spray water flow rates of the second stage attemperator, along with the final outlet steam temperatures of the superheat and reheat sections, are not significantly affected by a change in conduction resistance on any of the heat exchanger legs. These measurements are therefore also neglected.

Using a reciprocal analysis with a disturbance angle of five degrees, it is found that a linearly independent fault matrix does not require the flue gas exit temperature measurement and can be achieved by only considering the interstage steam temperature and spray water flow measurements. Furthermore, not all these measurements are needed. There are eight different combinations of these measurements that produce a linearly independent fault matrix. These combinations are shown by the shaded cells in Table 4-1. Note that each measurement in Table 4-1 represents all four corresponding measurements on legs A, B, C and D.

Table 4-1: Measurement combinations yielding a linearly independent conduction resistance fault matrix

Combination	**Measurements**							
1	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}
2	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}
3	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}
4	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}
5	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}
6	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}
7	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}
8	T_{stsh1o}	$\dot{m}_{att1}$	T_{stsh2i}	T_{stsh2o}	T_{stsh3i}	T_{strh1o}	$\dot{m}_{att3}$	T_{strh2i}

Legend		
T_{stsh1o}	-	SH1 steam outlet temperature
$\dot{m}_{att1}$	-	Fist stage attemperator spray water mass flow rate
T_{stsh2i}	-	SH2 steam inlet temperature
T_{stsh2o}	-	SH2 steam outlet temperature
T_{stsh3i}	-	SH3 steam inlet temperature
T_{strh1o}	-	RH1 steam outlet temperature
$\dot{m}_{att3}$	-	Third stage attemperator spray water mass flow rate
T_{strh2i}	-	RH2 steam inlet temperature

The index matrix resulting from the reciprocal analysis of the seventh combination in Table 4-1 is shown in Figure 4-12. A colour is assigned to each value in the matrix according to the colour scale

in Figure 4-12. Similar results are obtained for the remaining combinations and are included in Appendix E.

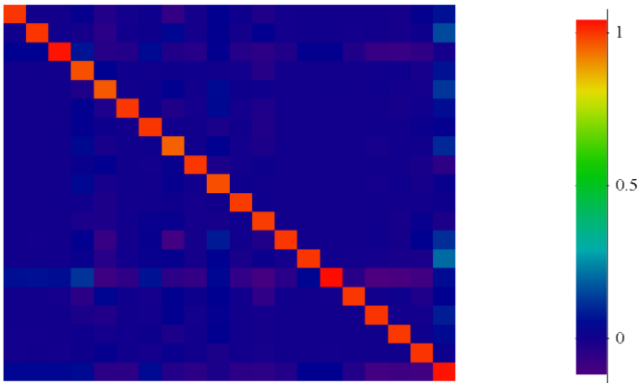


Figure 4-12: Reciprocal analysis with five degree disturbance angle for conduction resistance faults with steam temperature measurements only

The index matrix closely represents the identity matrix, indicating weak linear dependence. Therefore, by only considering 20 measurements, the conduction resistance of each leg can be individually diagnosed.

The analysis up to here only considered operation at fixed process boundary conditions. Changes in process boundary conditions that affect any of these 20 measurements will also cause changes in the conduction resistance diagnosis results, even if no change took place, resulting in misdiagnosis.

This happens because changes in process boundary conditions will cause the NOP to move. Adjusting the NOP according to these changes will correct the misdiagnosis. The NOP can be adjusted using design information, historical operating data or model simulation with the current process boundary conditions. Some difficulties arise with these approaches. Design information allows the NOP to be adjusted according to load. Any change in process boundary conditions that do not correspond to the design information at a given load will therefore cause this NOP to be incorrect. Using historical operating data also requires the process boundary condition changes to be aligned with that experienced during the historical period considered. Also, the historic operating period might contain faults and therefore lead to an incorrect NOP. Simulating a model with no faults for each new process boundary condition to obtain the NOP defeats the purpose of a data-driven monitoring method.

Another approach of dealing with changing process boundary conditions is to, instead of moving the NOP, include each process boundary condition change as a fault. Then the contribution of

these changes to the moving measurement point can be accounted for. Changes in the following process boundary conditions are included as faults:

1. Superheat steam inlet temperature
2. Total superheat steam mass flow rate
3. Reheat steam inlet temperature
4. Total reheat steam mass flow rate
5. Flue gas inlet temperature of SH2 leg A
6. Flue gas inlet temperature of SH2 leg B
7. Flue gas inlet temperature of SH2 leg C
8. Flue gas inlet temperature of SH2 leg D
9. Total flue gas mass flow rate

Including these process boundary condition changes as faults increases the number of faults to 29, meaning that the number of measurements now considered must be at least equal to this. The process boundary condition measurements provide nine additional measurements, totalling 29 (with those considered for the conduction resistance faults). However, a reciprocal analysis with a disturbance angle of only one degree does not show any resemblance to an identity matrix, indicating strong linear dependency.

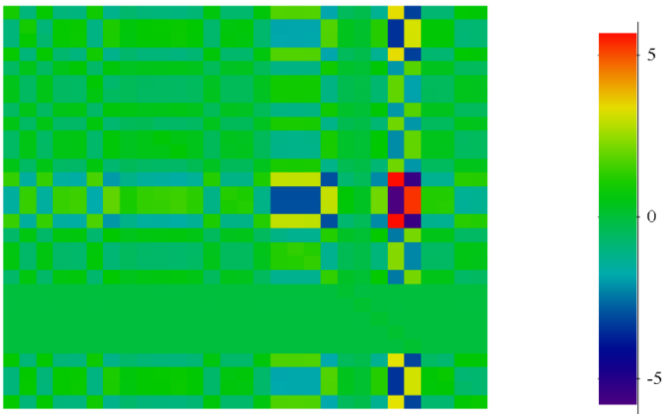


Figure 4-13: Reciprocal analysis with one degree disturbance angle for conduction resistance and process boundary condition faults

The measurements of spray water flow rates of the second stage attemperator and the final outlet steam temperatures of the superheat and reheat sections are also not significantly affected by a change in any process boundary condition. These measurements are therefore still not considered. Inclusion of the remaining 12 measurements of interstage steam temperature and spray water flow does not alleviate the linear dependency between these fault vectors. The

number of faults must therefore be decreased. If the flue gas inlet temperature is not considered as discretized, but rather unified, this reduces the number of faults to 26 and linear dependency is achieved by only considering 29 measurements. This includes the nine process boundary condition measurements and the 20 measurements required for linear independency of the conduction resistance fault matrix.

It is further observed that the four steam mass flow rate measurements at the final superheater outlet only responds significantly to changes in the total steam mass flow rate out of the superheat section. The response of these four measurements is also similar, being directly proportional to the changes in total steam mass flow rate. No additional information is therefore provided by considering these measurements separately and they are therefore summed to represent a single measurement of the total steam mass flow rate. The number of measurements is consequently reduced to be equal to the number of considered faults, being 26, without loss of linear independency. The reciprocal analysis for these 26 faults and 26 measurements are shown in Figure 4-14 for a disturbance angle of five degrees.

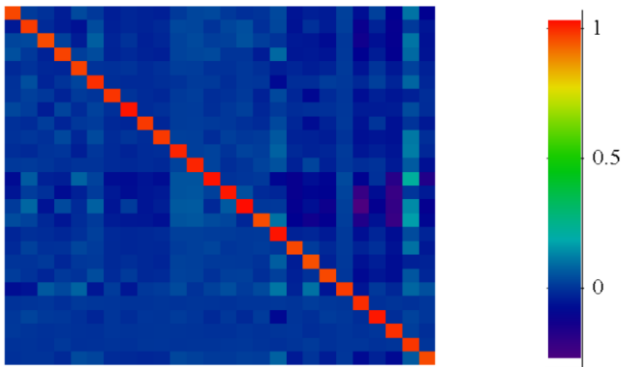


Figure 4-14: Reciprocal analysis with five degree disturbance angle for conduction resistance and process boundary condition faults with unified flue gas inlet temperature

Other faults that can be considered are localised cooling by soot blowing or air ingress, blockages on the steam and gas sides and steam leaks. These effects were imposed on the model and it was found that a linearly independent fault matrix cannot be obtained for these faults when conduction resistance faults are included. These faults will therefore not be included in the fault matrix. It should be noted that if blockages or localised cooling are present, their effect on the measurement space will cause changes in the diagnosis results of the faults included in the fault matrix.

Evaluation of process nonlinearity

The nonlinearity of the process measurements' response to faults can influence fault vector directions according to the fault magnitude, the load and the presence and magnitude of other faults. If the nonlinearity is severe, the fault matrix should be constructed with several simulations for each fault at different load and fault conditions and the fault matrix should ideally be updated with changes in load and diagnosis results during monitoring. Otherwise, for relatively linear response, only one training example per fault type can be motivated. An evaluation is therefore done to quantify this nonlinearity by comparing fault vectors for a range of fault magnitudes and loads. Extensive evaluation is required to compare the effect of all possible fault combinations on the direction of a specific fault vector and falls outside of the scope of this study. A condensed evaluation will however be presented to see how the simultaneous increase in all conduction resistance faults influence the fault vector for the conduction resistance of a single leg at a constant magnitude. The effect of specific structured changes in process boundary condition faults on fault vector directions is also encapsulated in the evaluation of load variations.

To relate the fault magnitude to the distance covered along the direction of its fault vector requires knowledge about the rate at which these two quantities covary. An evaluation is therefore also conducted to quantify this relation.

Effect of fault magnitude on fault vectors

Quantification of nonlinearity is purpose specific and dependant on the range of evaluation. The approach used here first obtains a fault vector for a specified fault magnitude to serve as a baseline. The fault vectors obtained for different fault magnitudes are then compared to this baseline fault vector. These vectors are compared by evaluating the angle between their directions. A conceptual example is given in Figure 4-15. Here it can be seen how a fault causing a nonlinear process measurement response leads to a deviation from the baseline fault vector direction by some angle.

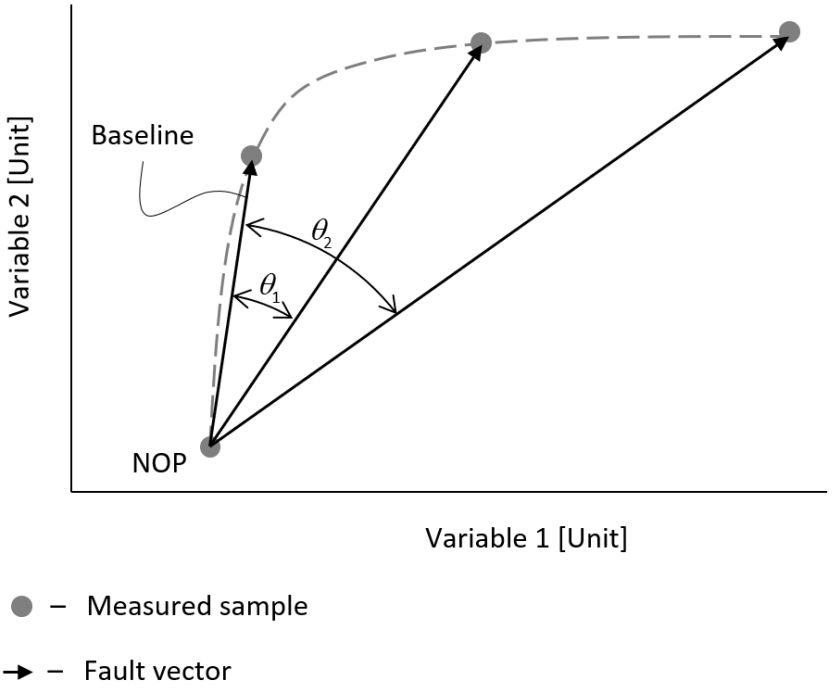


Figure 4-15: Nonlinearities causing the direction of fault vectors to change by some angle

The fault vectors corresponding to different fault magnitudes are compared by imposing five incremental increases on the magnitude of each fault. The increment size for each fault is shown in Table 4-2 and a number is assigned to each fault for ease of reference.

Table 4-2: Fault magnitude increments and number assignment

Fault number	Fault type and location		Increment size
0	Conduction thermal resistance	SH1A	0.4 [K/MW]
1		SH1B	
2		SH1C	
3		SH1D	
4		SH2A	
5		SH2B	
6		SH2C	
7		SH2D	
8		SH3A	
9		SH3B	
10		SH3C	
11		SH3D	
12		RH1A	

13		RH1B	
14		RH1C	
15		RH1D	
16		RH2A	
17		RH2B	
18		RH2C	
19		RH2D	
20	Superheat steam inlet temperature		0.9 [K]
21	Superheat steam outlet mass flow rate		5 [kg/s]
22	Reheat steam inlet temperature		5 [K]
23	Reheat steam outlet mass flow rate		5 [kg/s]
24	Flue gas total mass flow rate		5 [kg/s]
25	Flue gas inlet temperature		5 [K]

Figure 4-16 shows the propagation of the angle between the fault vectors for each fault as the imposed deviation is incremented from the baseline magnitude.

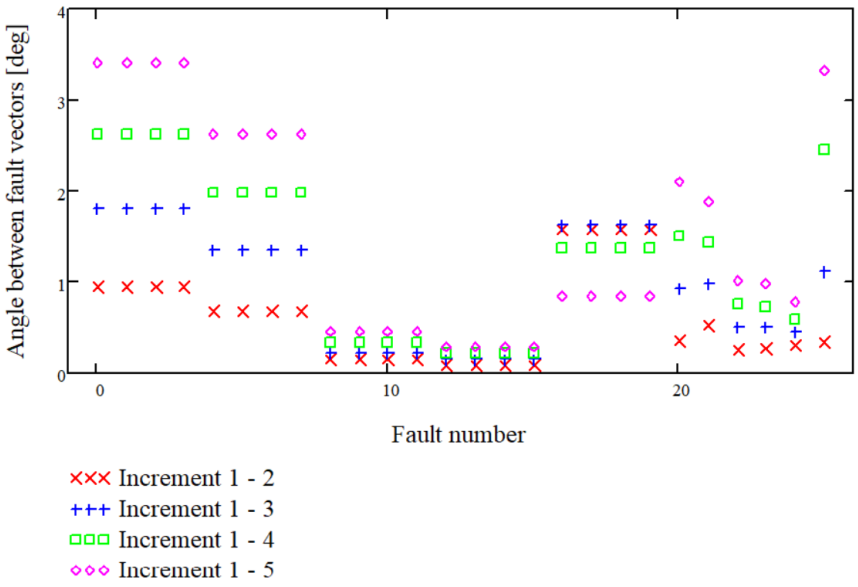


Figure 4-16: Angle of fault vector deviation by fault magnitude

The deviation in fault vector direction caused by varying fault magnitudes are within four degrees for the range of fault magnitudes tested. These angles being nonzero indicate that the movement

traced through the measurement space by increasing fault magnitudes is nonlinear. The fault vector directions for conduction resistance of SH3 and RH1 however seem to be less sensitive to fault magnitude for the range tested, indicating a more linear movement. It is also noted that the deviation angles for the conduction resistance faults of RH2 reveal a decrease with fault magnitude following an initial increase, revealing the presence of an inflection point. This indicates that, according to the model equations, the measurements are at least a third-degree function of the conduction resistance of RH2's legs.

To get a visual grasp on this multidimensional nonlinear movement, PCA (see Section 3.2) is used to reduce the major directions of the fault vector propagation onto two dimensions. For each fault type, the dataset containing the fault vectors corresponding to the different fault magnitudes are normalized and made unitless. PCA is applied to each dataset to identify the first two principal components which contain the most variation in the dataset. The datasets are then transformed onto the principal component subspace and plotted on the first two PCs. These plots are shown for conduction resistance faults on legs A of SH1, RH1 and RH2 in Figure 4-17. The first two PCs plotted on equal scales in Figure 4-17 fall short of only 1×10^{-7} % for SH1, 4×10^{-9} % for RH1 and 1×10^{-6} % for RH2 of the total variation in each dataset, allowing virtually all the movement to be visualised on the plots. Similar plots for the remaining faults are included in Appendix F. Note that the PCs are unitless.

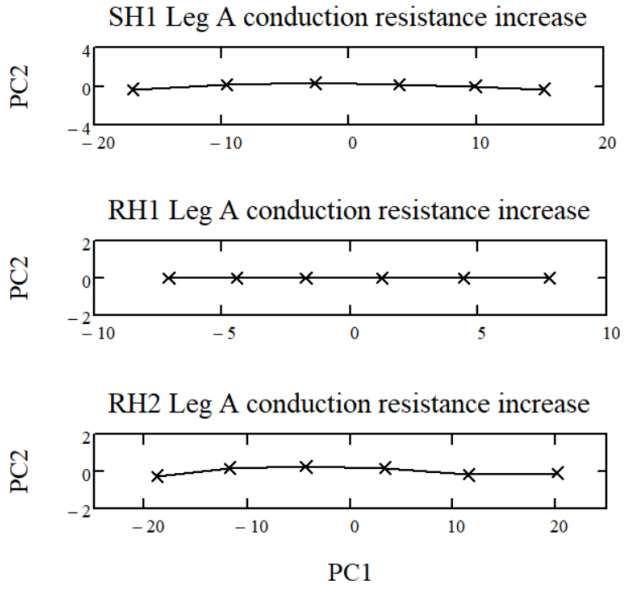


Figure 4-17: Nonlinear movement caused by decreasing conductivity on leg A of SH1, RH1 and RH2

Figure 4-17 substantiates the interpretation of Figure 4-16 as this visualisation reveals the fault vector propagations to be more linear for RH1 than for SH1 and RH2. The third-degree movement posited for RH2 is observed.

The extent of these nonlinearities was deemed to be small enough to allow the use of a single fault vector, irrespective of the fault magnitude, without adversely affecting the results. Not making this simplifying assumption will necessitate a continuous update of the fault matrix to accompany any change in the fault diagnosis. This would require more extensive training to generate fault vectors corresponding to a range of fault magnitudes. As diagnosis results change, the fault matrix would need to be replaced with one developed for fault magnitudes that match the new diagnosis results.

Effect of load variations on fault vectors

The fault vectors for conduction thermal resistance and process boundary condition changes are compared for different loads. The fault magnitudes are fixed at a single increment size given in Table 4-2.

Load changes are simulated by simultaneously adjusting all load dependent process boundary conditions. This information is obtained from the boiler design documentation where it is listed for three different load cases. The load cases correspond to 70%, 97% and 100% of the boiler maximum continuous rating (BMCR). Figure 4-18 presents the angles between the respective fault vectors at the different loads.

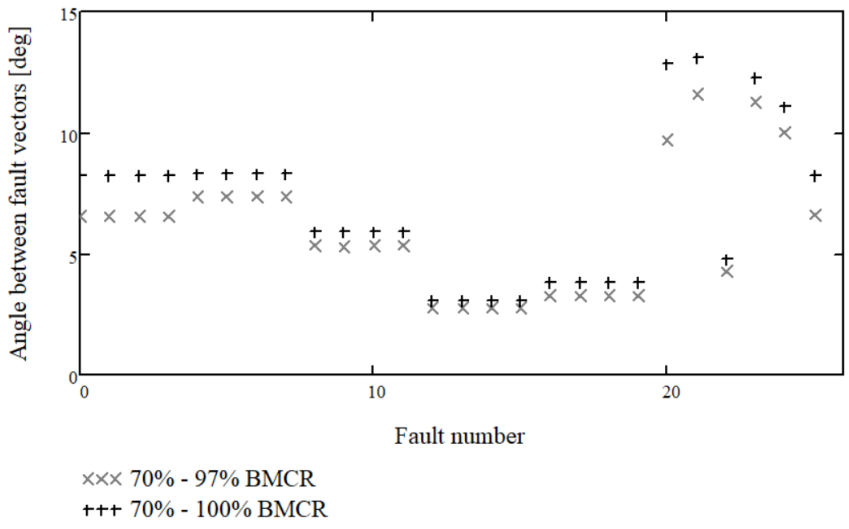


Figure 4-18: Angle of fault vector deviation caused by load variations

The changes in the fault vector directions caused by load changes are considerable. A fault matrix is therefore created for each of these load cases. Interpolation between these fault matrices is used in application to match the actual load.

Effect of other faults on fault vectors

The direction of a fault vector can be influenced by the presence of other faults. The effects of the simultaneous increase in conduction resistance of all legs at fixed process boundary conditions on the fault vector for the conduction resistance of one leg is evaluated in this section. A 0.4 [K/MW] increase is imposed in the conduction resistance of SH3 leg A while simultaneously imposing an incremental increase of ± 0.02 [K/MW] in that of all other heat exchanger legs. The propagation of the angle between the original fault vector and those resulting from incremental increase in other faults is plotted in Figure 4-19.

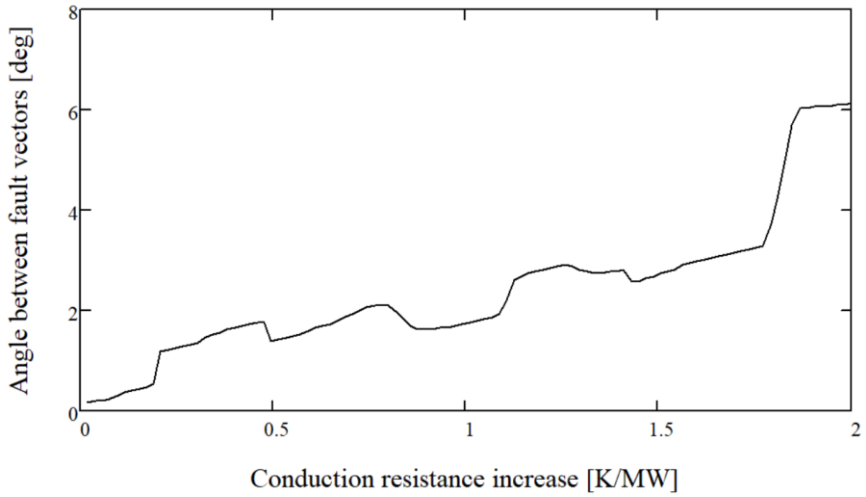


Figure 4-19: SH3 leg A fault vector deviation angle for increasing conduction resistance of all other legs

The nonzero angles indicate that the fault vector direction is influenced by the presence of other faults. Continuous changes in the conduction resistance of other legs appear to have a slight but sporadic influence on the fault vector direction of a single leg. The maximum deviation is within 7 degrees for the range tested. The influence of other faults on a fault vector will therefore be neglected. This allows the need for updating the fault matrix with every change in fault diagnosis to be averted.

Fault index to fault magnitude relation

The same simulation cases presented in Figure 4-16 are used in this section, but instead of looking at the difference in the angles between fault vectors, the relation between the fault vector magnitude and the magnitude of the deviation imposed on the fault variable is evaluated. For conduction resistance faults, only one leg is considered per heat exchanger stage as a similar

response is obtained for all legs. Figure 4-20 shows this relation for one leg from each heat exchanger stage.

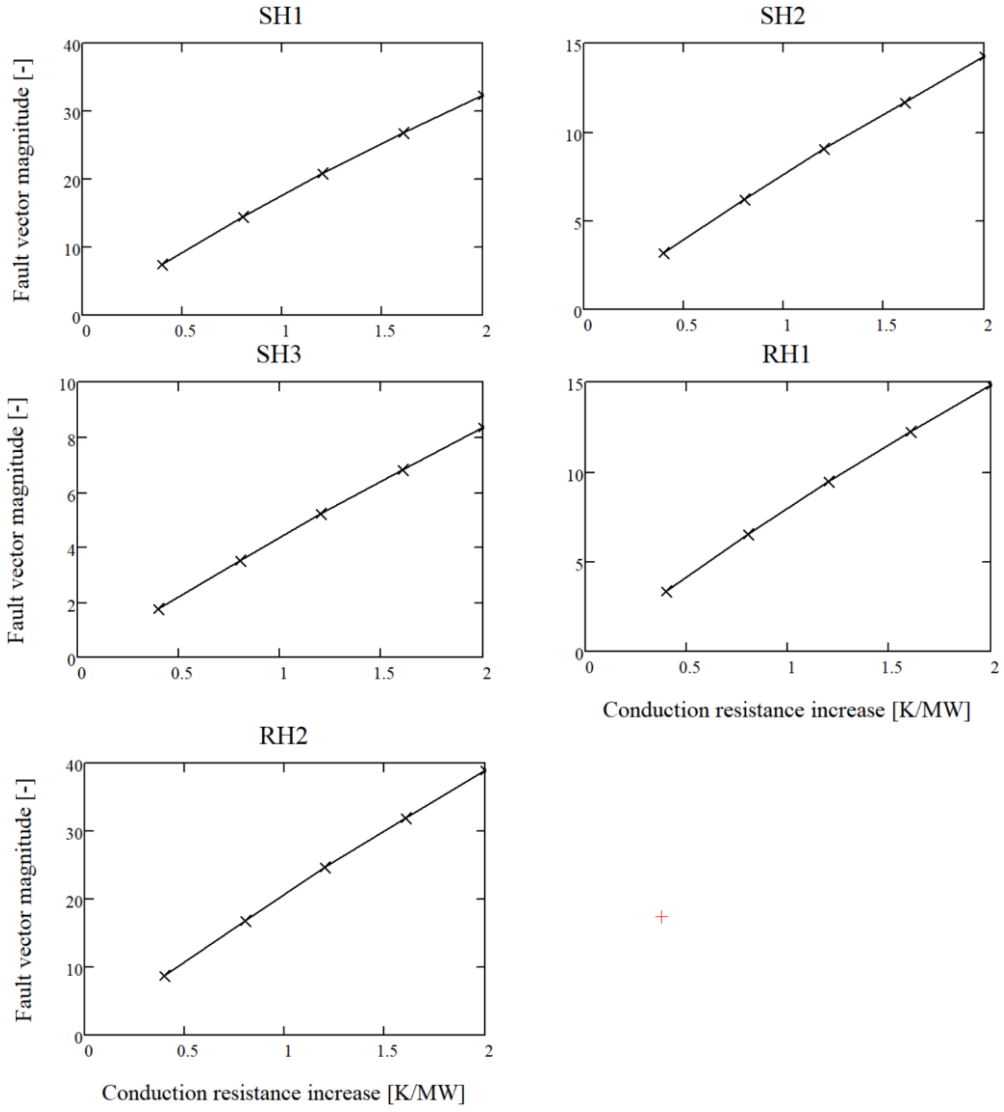


Figure 4-20: Distance of deviation from the NOP in relation to fault magnitude

The relations for process boundary condition faults are shown in Figure 4-21.

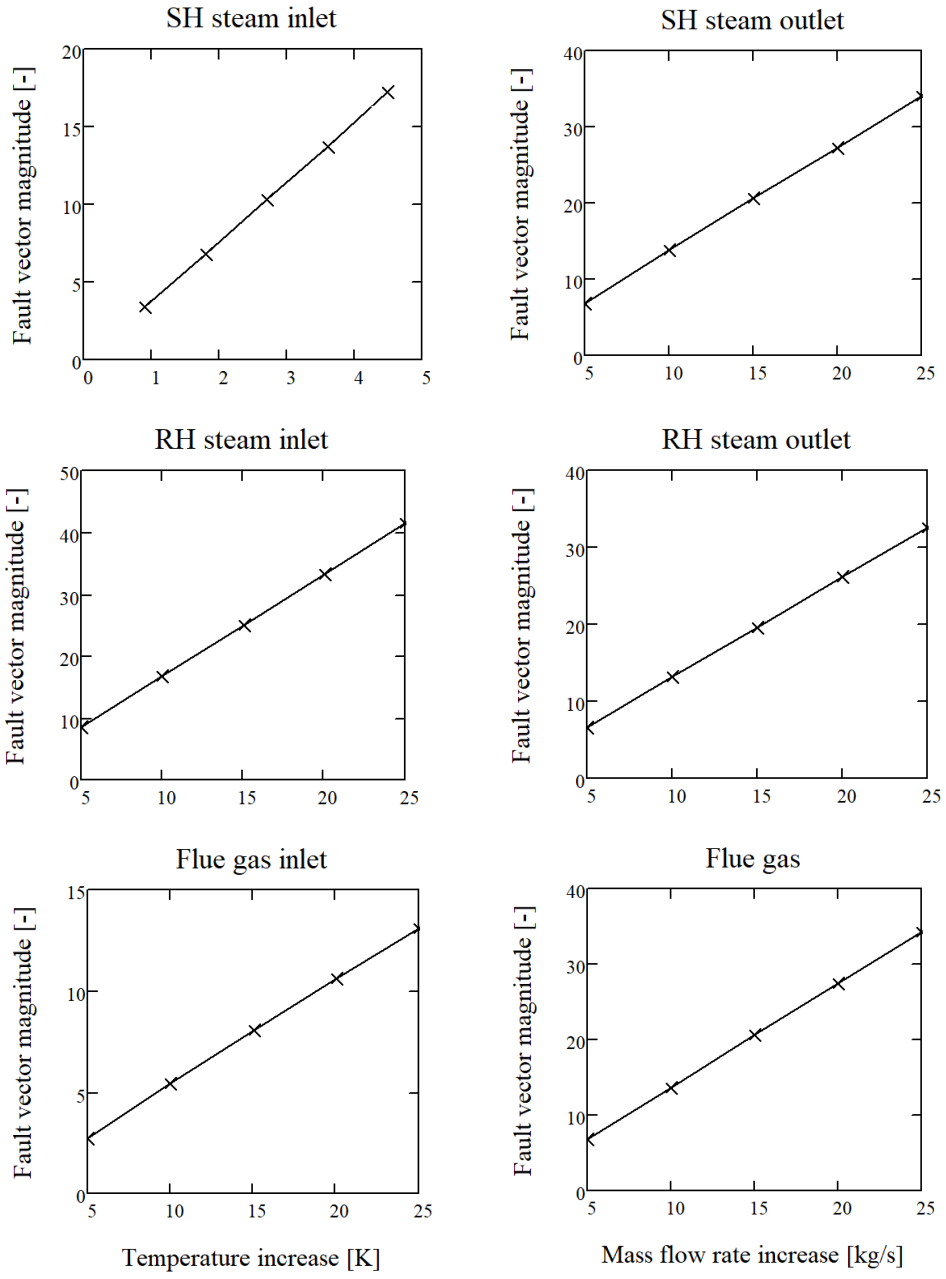


Figure 4-21: Process boundary condition fault vector to fault magnitude relation

A relatively linear relationship is observed between the fault variable deviation to the fault vector magnitude for both conduction resistance and process boundary condition faults. This allows the

absolute value of the fault variables to be obtained from the fault index by using the slopes for the respective faults in Figure 4-20 and Figure 4-21.

Summary

The faults selected for composing the fault matrix include changes in conduction resistance of each leg and process boundary condition changes. The measurements considered are all process boundary condition measurements and employs combination seven of the interstage steam temperature and spray water flow rate measurements in Table 4-1.

The effects of load changes on the fault vectors will be accounted for by simulating faults for a range of loads and interpolating between the resulting fault libraries. The effects of other faults on the fault vector directions will be neglected. The effects of different fault magnitudes on fault vectors will be neglected. Because of these simplifying assumptions, only one training example is required for each fault at a specified load. Also, the need for updating the fault matrix according to diagnosis results is averted.

A choice must however be made about the magnitude of the faults to be imposed on the model when generating the fault vectors. During large deviations from normal operation more emphasis is placed on the identification of the faulty variable than its quantification. The accuracy of the absolute value of a monitored variable might be of greater interest for small deviations in fault variables. The magnitude of the imposed faults during training are therefore chosen to be small and a single increment size from Table 4-2 is used.

5. Results

Actual sensor data, retrieved from the distributed control system (DCS) of the case study boiler, is used to compare the soft sensor and data-driven monitoring methods. The recorded period contains a load change and a soot blowing event. Soot blowing is the blowing of some medium onto the tube outer surfaces to remove ash (soot) build-up. The case study boiler uses steam as the soot blowing medium. The boiler is equipped with tube leak detectors (TLD) which are sensitive to the range of acoustic waves in gasses produced by passing steam. The TLDs therefore also respond to soot blowing with steam, giving an indication of when soot blowing is active.

The reason for using data containing a soot blowing event is that it is expected that the heat exchanger performance will respond to such an event in a somewhat predictable way. The test will therefore reveal whether a change in the heat exchanger performance can be picked up due to the soot blowing activity. Inclusion of a load change in the test data also allows the response of results to load changes to be observed.

The sensor data covers a 24-hour operating period, recorded at 5-minute intervals, resulting in 288 timesteps. The boiler is equipped with a total of 76 soot blowers in the convective pass, about 15 blowers per heat exchanger stage. Information about the specific blower being active is not available on the DCS, and only a general indication given by the TLDs is available. The load and TLD activity are plotted in Figure 5-1 against the recorded time, along with the periods concluded to be active soot blowing indicated by darkened backgrounds.

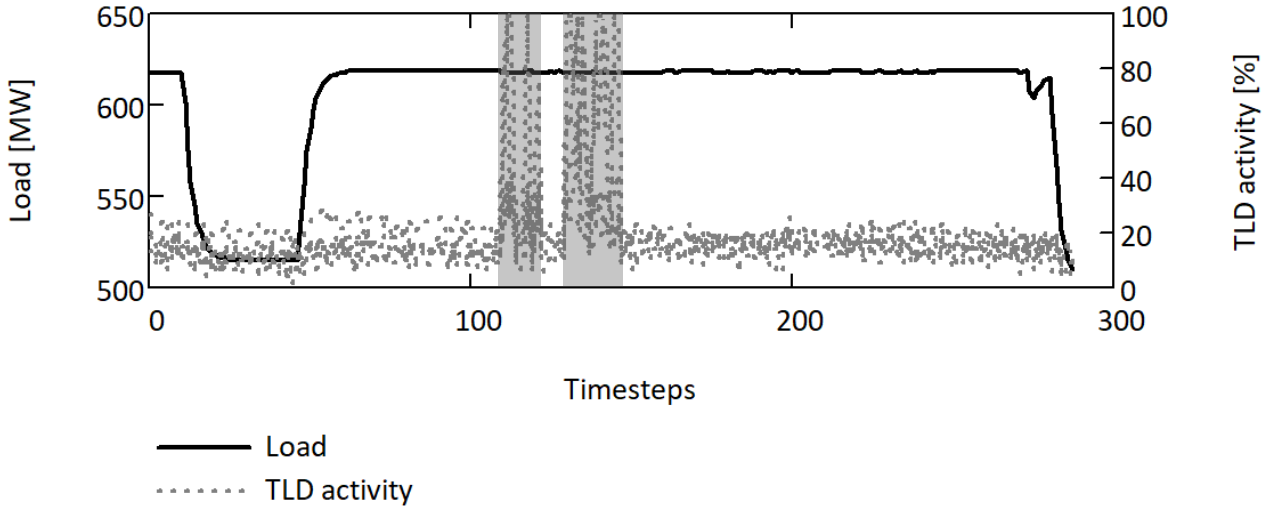


Figure 5-1: Load and TLD activity plotted against recording time

5.1 Soft sensor results

The soft sensor performs a simulation for each timestep, using the measured data as inputs. The computation time required for each simulation ranges from 5 to 15 minutes. All fluid variables throughout the system is then estimated according to the discretization of the model, representing the average values in and surrounding each leg.

The heat transfer rate to each leg as calculated by the soft sensor is shown in Figure 5-2 for the recorded period.

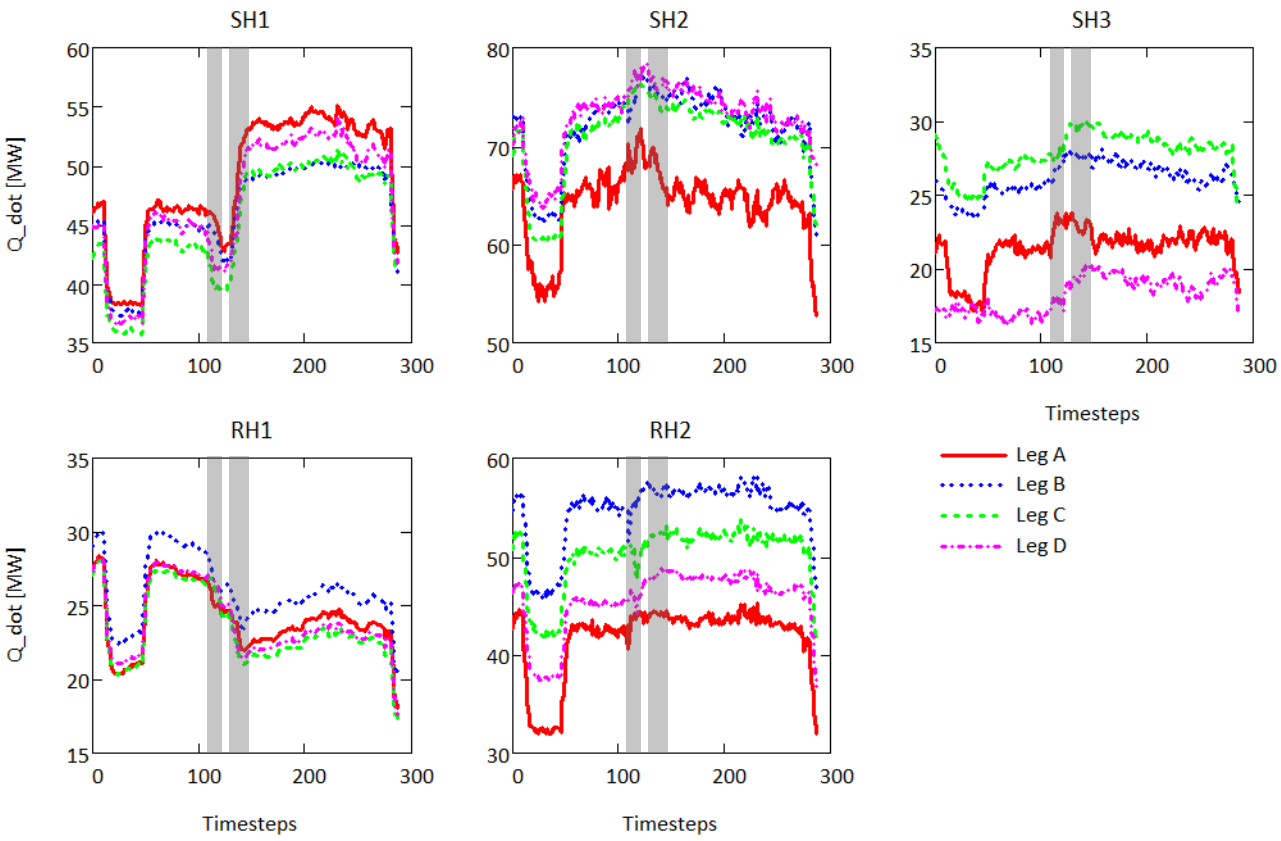


Figure 5-2: Heat transfer rate of individual heat exchanger legs

A decrease in heat transfer rates is seen with decreasing load. This is expected because of the lower steam and flue gas mass flow rates at lower loads. Soot blowing leads to an increase in the heat transfer rate of all heat exchanger stages, except for RH1 which experiences a decrease in heat transfer rate. This occurs even though the conduction thermal resistance of RH1 clearly decreases during soot blowing as shown in Figure 5-8. To note here is that the heat transfer rate is not an accurate indication of fouling as it is also influenced by other process conditions.

Through its effect on flow resistance and spray water flow rates, the rate of heat transfer to the different legs influences the distribution of steam flow through these legs. The derived ratio of the steam mass flow rate through each leg to the average of the stage is plotted in Figure 5-3.

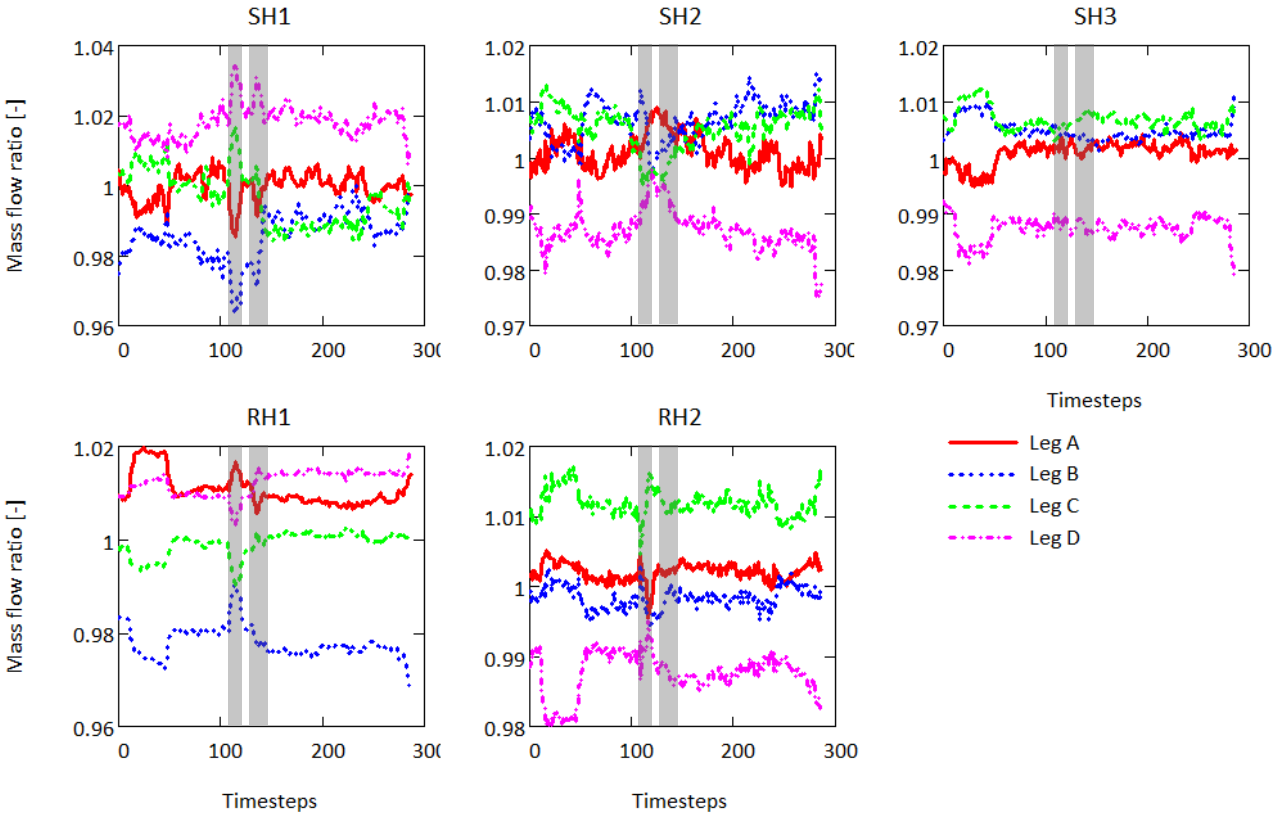


Figure 5-3: Steam mass flow rate as a ratio to the average

The derived steam mass flow rate in the sections of common headers connecting different legs responding to pressure imbalances are shown in Figure 5-4. It can be seen that the steam flows in headers connecting different leg pairs are small in relation to the total steam flow through these legs. Consequently, the average temperatures at the inlet or outlet of a leg correspond very closely to the temperature at the corresponding measurement location on the interconnecting piping.

Nonuniform process conditions also lead to an uneven flow distribution in the flue gas flow path. The soft sensor model discretizes the flue gas flow into four parallel streams with flow connections between each stage (Figure 4-4). The calculated flue gas mass flow rate through each leg as a ratio to the average is plotted in Figure 5-5 for each stage.

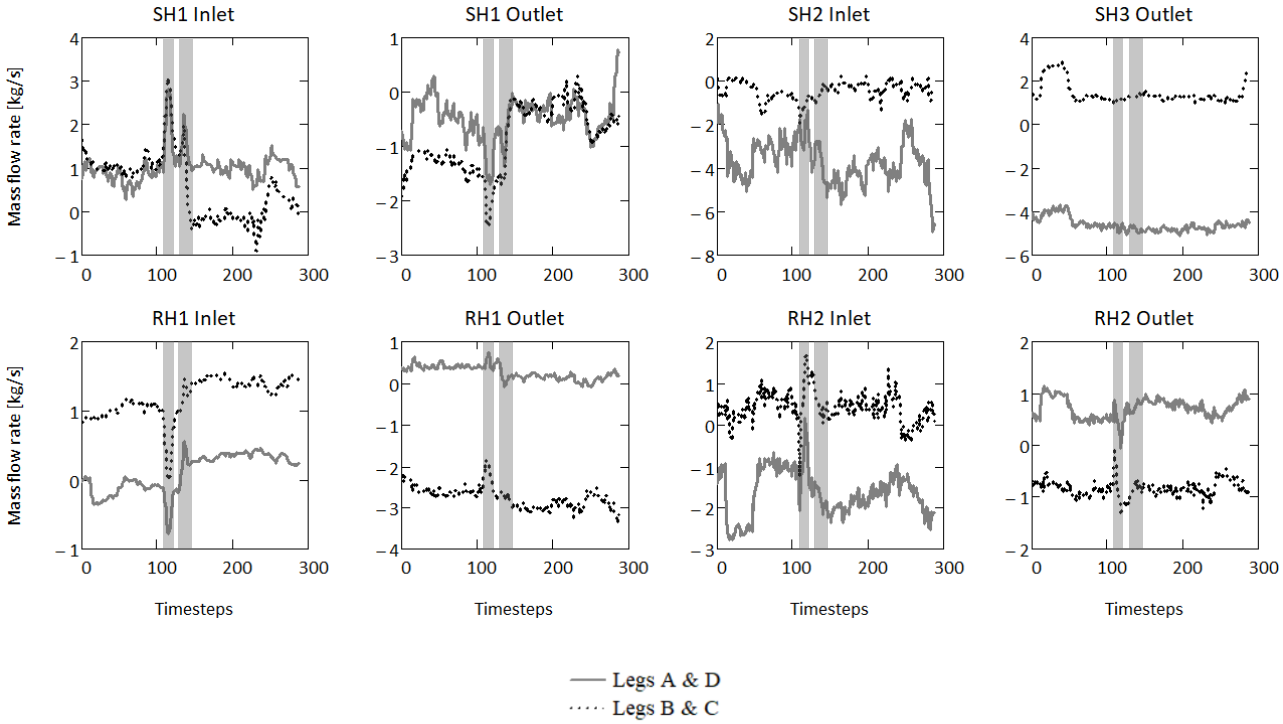


Figure 5-4: Steam flow between connected legs in common headers

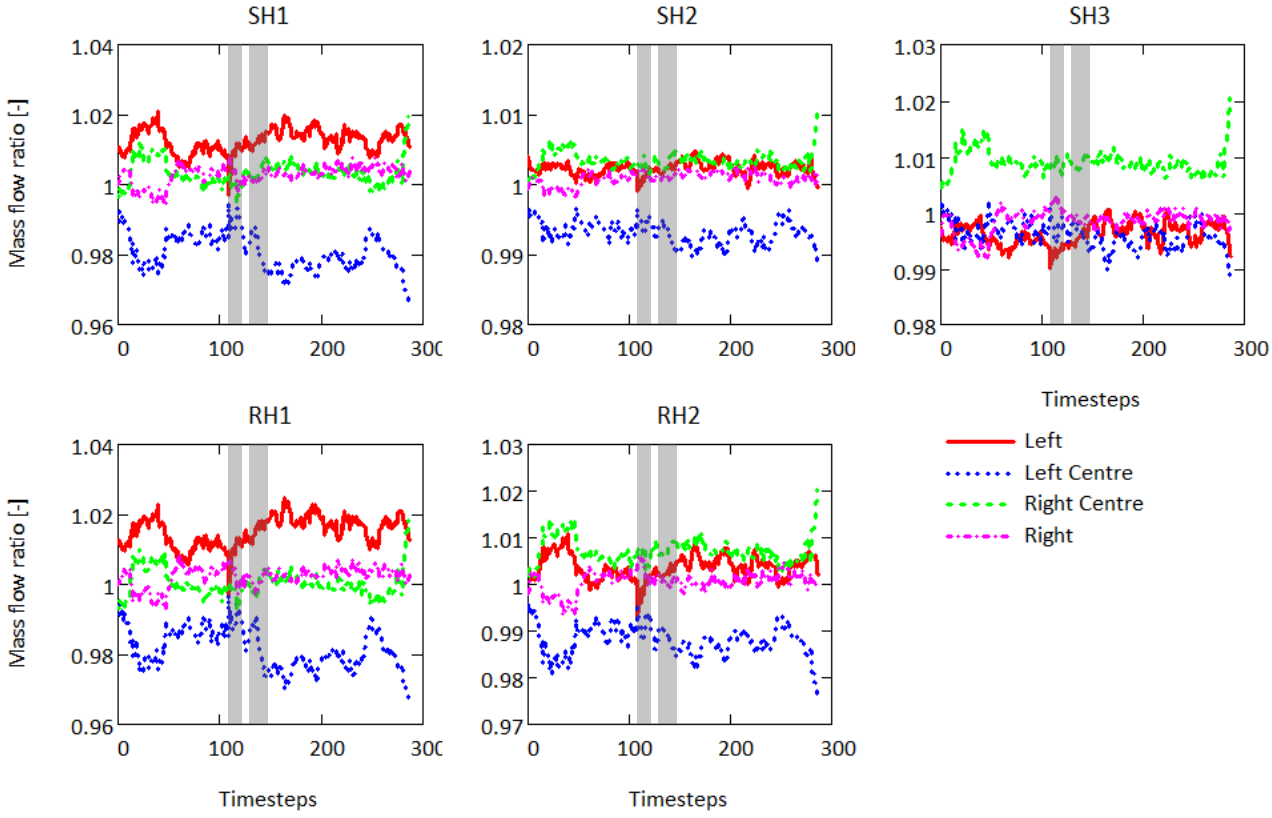


Figure 5-5: Flue gas mass flow rate as a ratio to the average

The flue gas flowing between the two left streams, the two centre streams and the two right streams in-between each of the heat exchanger stages as calculated with the soft sensor is shown in Figure 5-6.

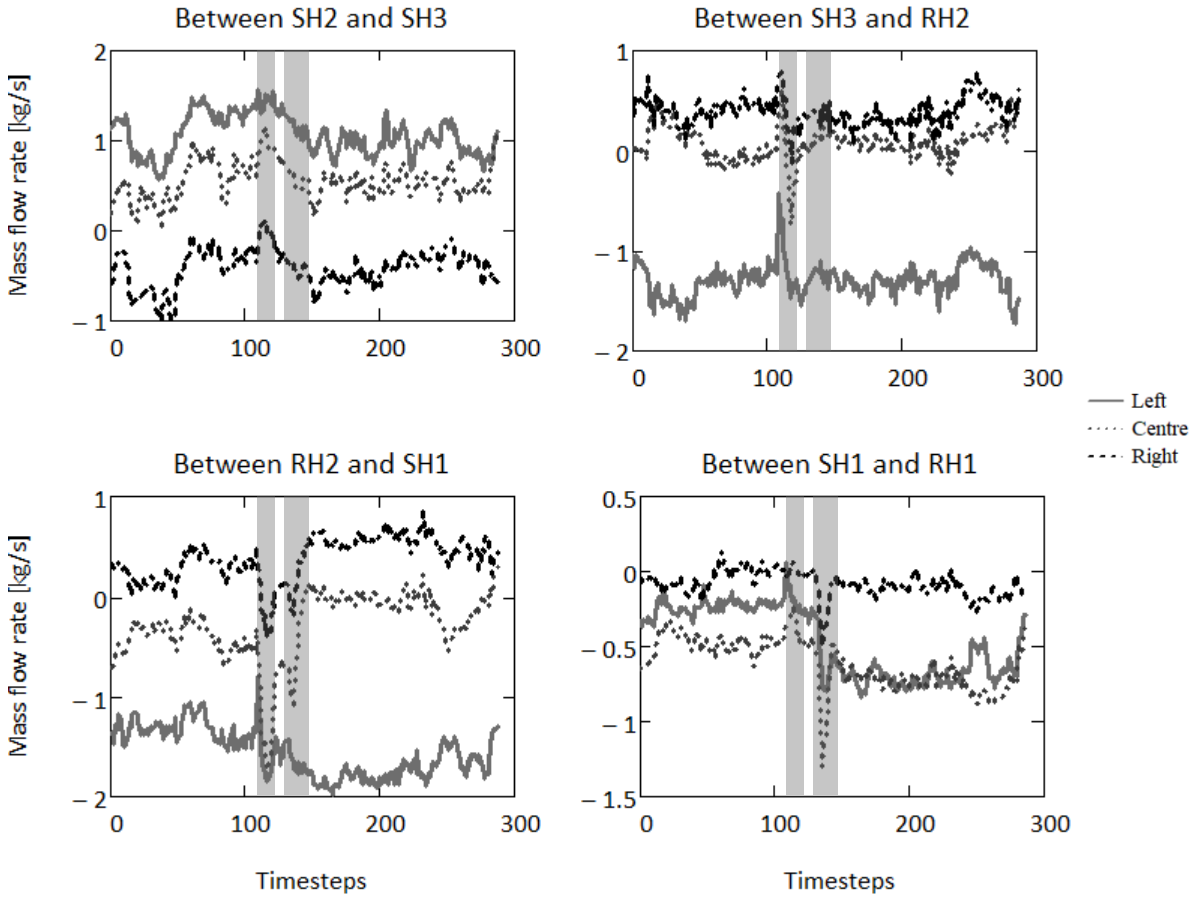


Figure 5-6: Interstage flue gas flow between the four flow streams

In Figure 5-3 to Figure 5-6 it is seen that both steam and flue gas flow are not uniformly distributed amongst the legs of a stage. The extent of maldistribution is however not severe and will have minimal impact on the average tube metal temperature and fly ash erosion rates. Significant changes in the flow distributions are triggered by the load change and soot blowing events. This is probably because the change in heat transfer rate during these events impacts the temperature distribution and consequently the pressure drops by density differences. The flue gas flow distribution seems to be correlated with flue gas temperature distribution downstream of a heat exchanger stage shown in Figure 5-7.

Significant differences in steam temperatures amongst legs within a heat exchanger stage are seen in the plant measurements. These temperature differences are caused by significant differences in the heat transfer rates of different legs (Figure 5-2). The differences in heat transfer rates are caused by and in turn results in uneven flue gas temperatures profiles across heat exchanger

stages. The flue gas temperatures entering each stage as estimated with the soft sensor is shown in Figure 5-7 for the respective legs.

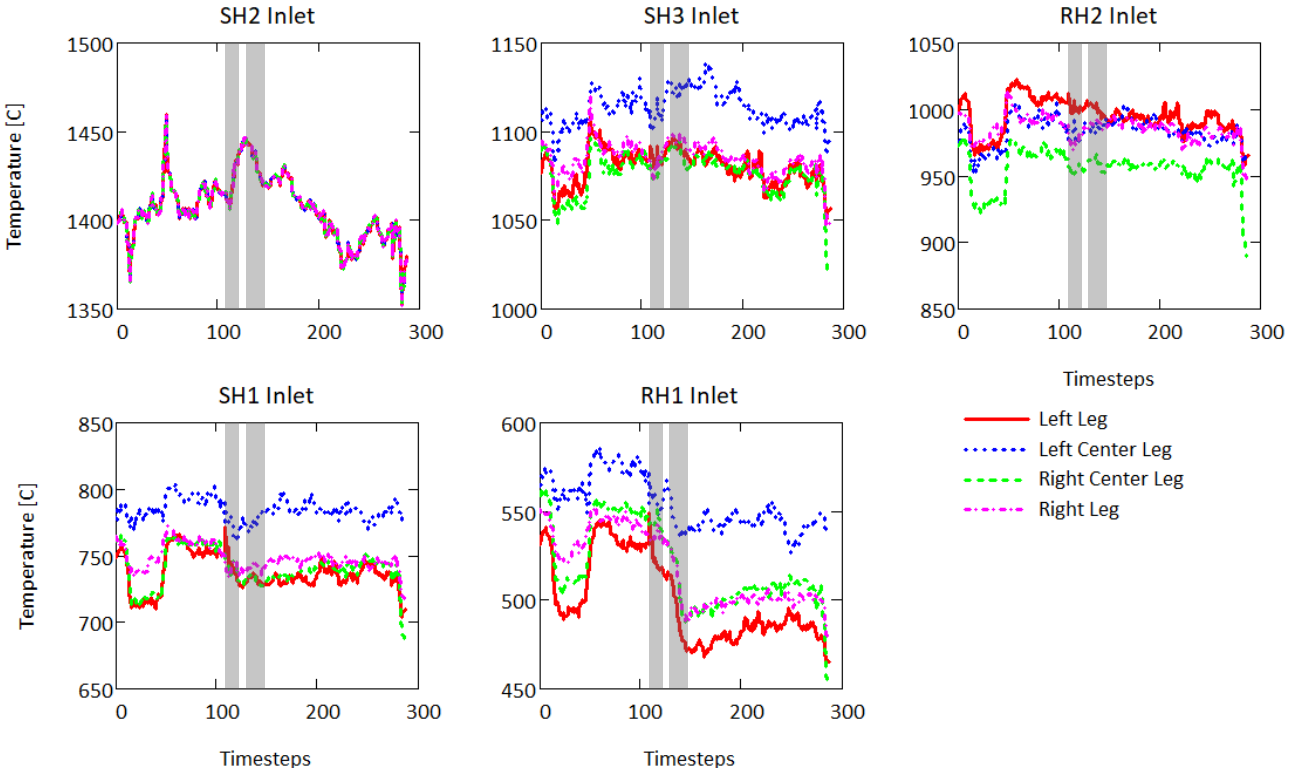


Figure 5-7: Flue gas temperature

The flue gas temperatures into SH2 and SH3 are relatively stable during the load change and soot blowing, with the temperatures into the heat exchanger stages further downstream being more sensitive to these events. The variations in the calculated flue gas inlet temperature are potentially a result of process controls acting on upstream processes not included in this model.

The model is configured with a single flue gas inlet source. The temperature of this source is calculated to achieve the flue gas temperature of the single measurement at the economiser inlet. Because only one temperature sensor is located at the economiser inlet, the soft sensor can only calculate a uniform flue gas inlet temperature profile (SH2 Inlet in Figure 5-7). The temperature drops over each leg are superimposed on this uniform profile, resulting in a flue gas temperature profile discretized to the level of each leg between the heat exchanger stages up to the outlet of RH1. The mass-based average temperature of the four flow streams at the outlet of RH1 equals the single temperature measurement at the economiser inlet. Attempting to calculate the flue gas inlet temperature discretized for each leg results in a situation of more unknowns than equations. The unknowns required can be provided by forcing a uniform temperature profile at the economiser inlet, where the temperature of each flow stream out of RH1 is equal to the

economiser inlet measurement. The temperature drops over each leg are then superimposed on this uniform profile up to the flue gas inlet, resulting in discretized flue gas temperatures at the inlet of SH2. It is therefore evident that access to discretized temperature measurements for the four flow streams at the economiser inlet would allow the discretized flue gas inlet temperature to be more accurately determined.

The soft sensor accounts for thermal resistance to radiation and convection between the flue gas and the outer surface of the tube or fouling layer and assumes the remainder to be conduction through a layer of fouling or oxide. It can therefore distinguish between the different kinds of thermal resistances. The various components of the calculated thermal resistance are shown in Figure 5-8 to Figure 5-10.

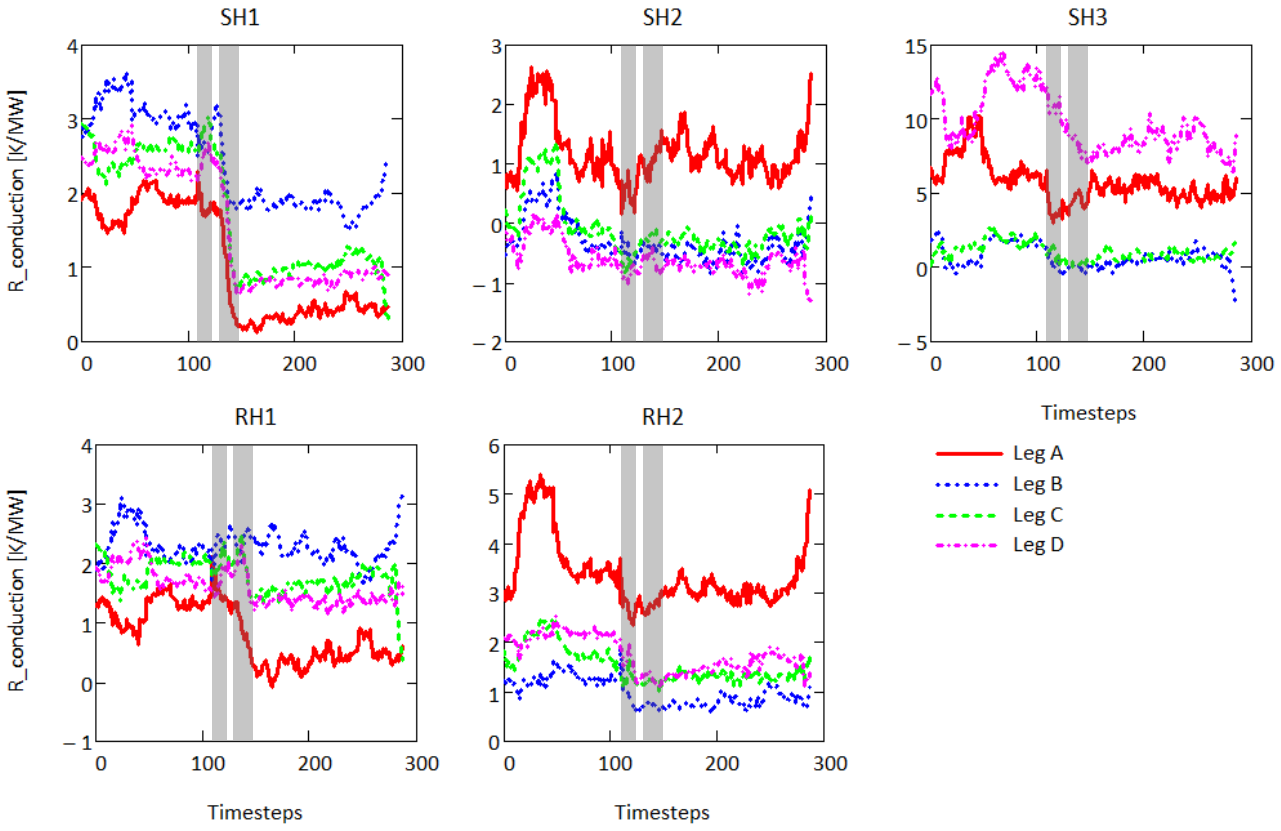


Figure 5-8: Conduction thermal resistance

Conduction resistance of less than zero indicates that the actual overall thermal resistance is less than the value used in the design. This can be caused by an overestimation of the baseline fouling resistance or other thermal resistances in the design.

As expected, soot blowing causes a decrease in the conduction thermal resistance of all legs (Figure 5-8). The conduction resistance of some legs also responds to the load changes. Decreasing conduction thermal resistance during load changes may be caused by thermal expansion or

contraction of tube banks resulting in ash falling from tubes (Díez et al. 2005). However, a sudden increase in conduction resistance cannot be substantiated and is most probably a result of modelling uncertainty. Changing process conditions influence the convection and radiation thermal resistance in a predictable manner as determined by the respective empirical correlations. If the changes in these resistances described by the model are for some reason not physically realistic, this error will influence the calculated response in the conduction thermal resistance results via equation (4.3).

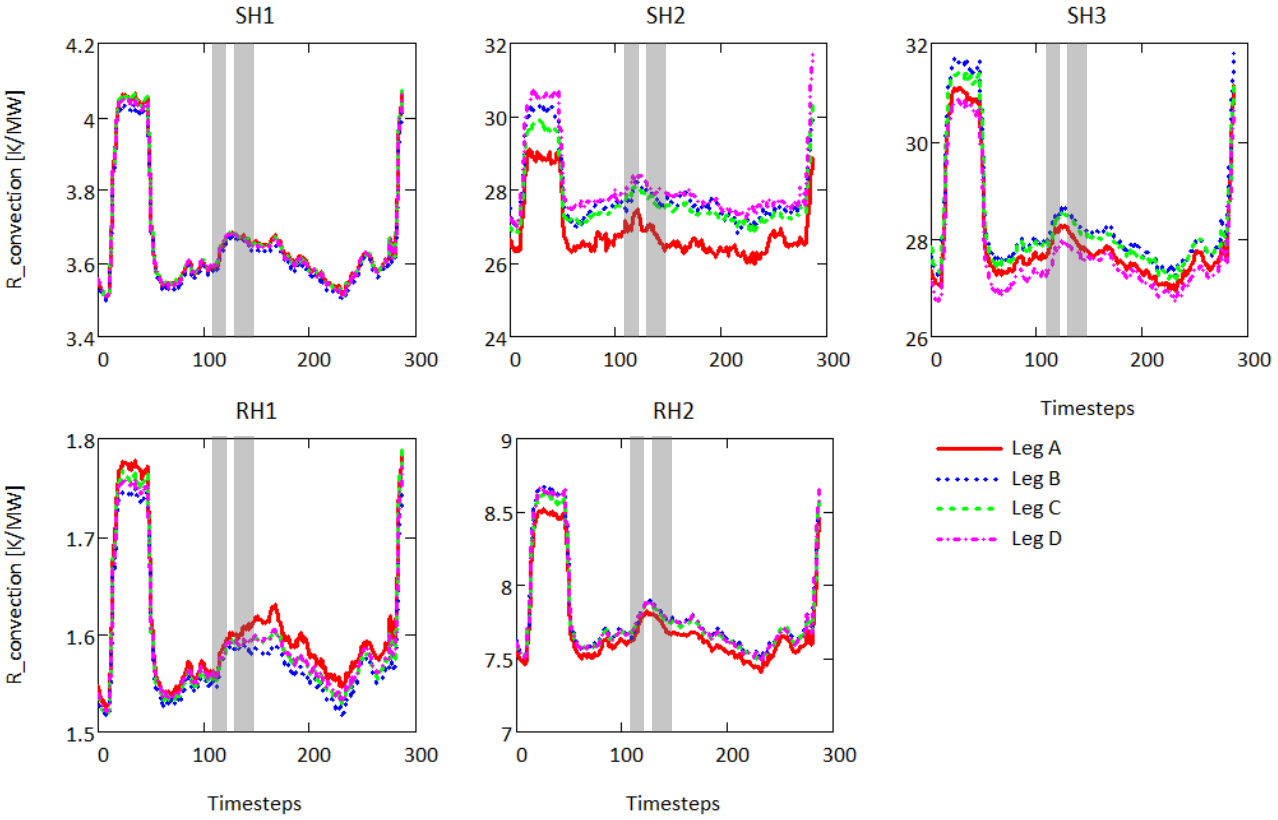


Figure 5-9: External convection thermal resistance

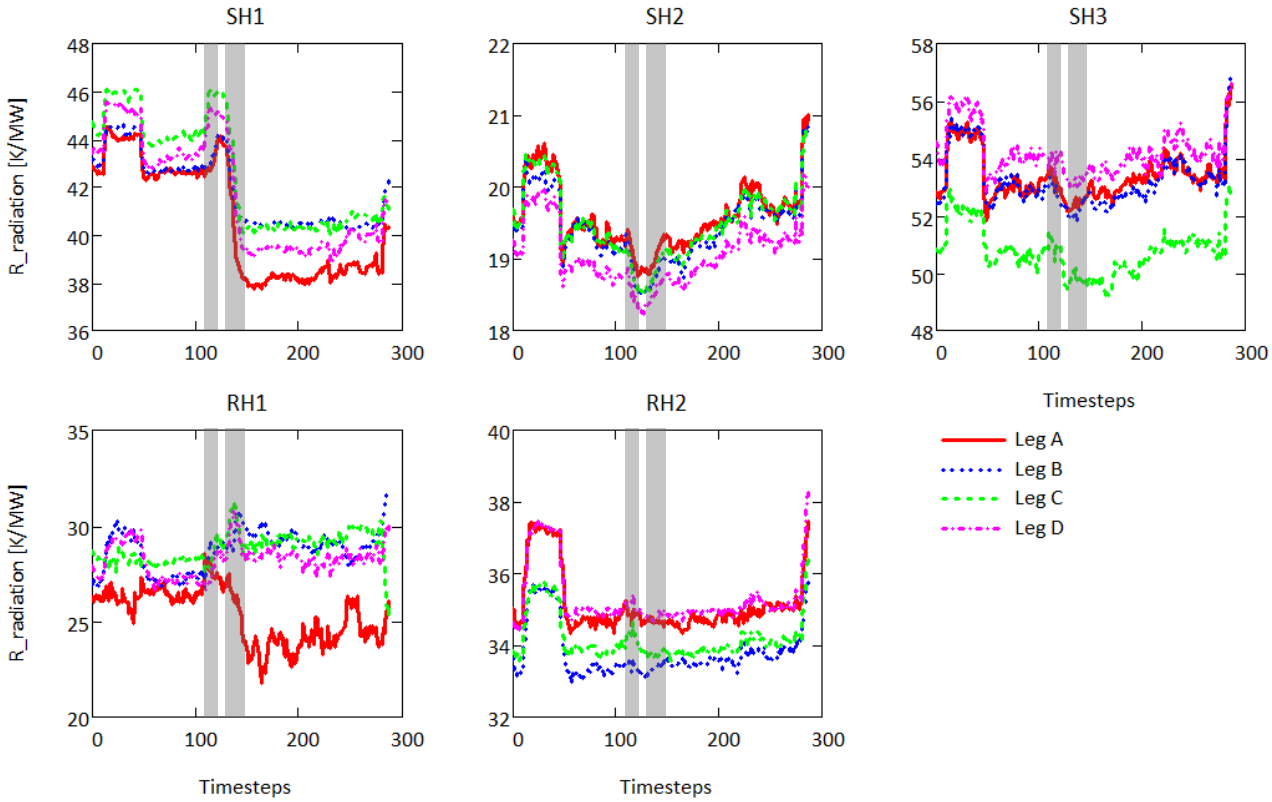


Figure 5-10: External radiation thermal resistance

The difference between the radiation thermal resistance amongst the legs (Figure 5-10) are more significant than the difference in convection resistance (Figure 5-9). The radiation heat transfer resistance depends on the temperature difference to the fourth power as per equation (3.20) and the flue gas temperature maldistribution across a heat exchanger stage therefore has a significant effect on the radiation thermal resistance of the different legs. The convection thermal resistance depends on flow rate and other flue gas properties to powers less than one as per equation (4.2) and is relatively similar amongst the different legs of a stage.

The temperature difference between the fluids effects a heat transfer rate governed by the heat transfer resistance. The log mean temperature difference for each leg is shown in Figure 5-11.

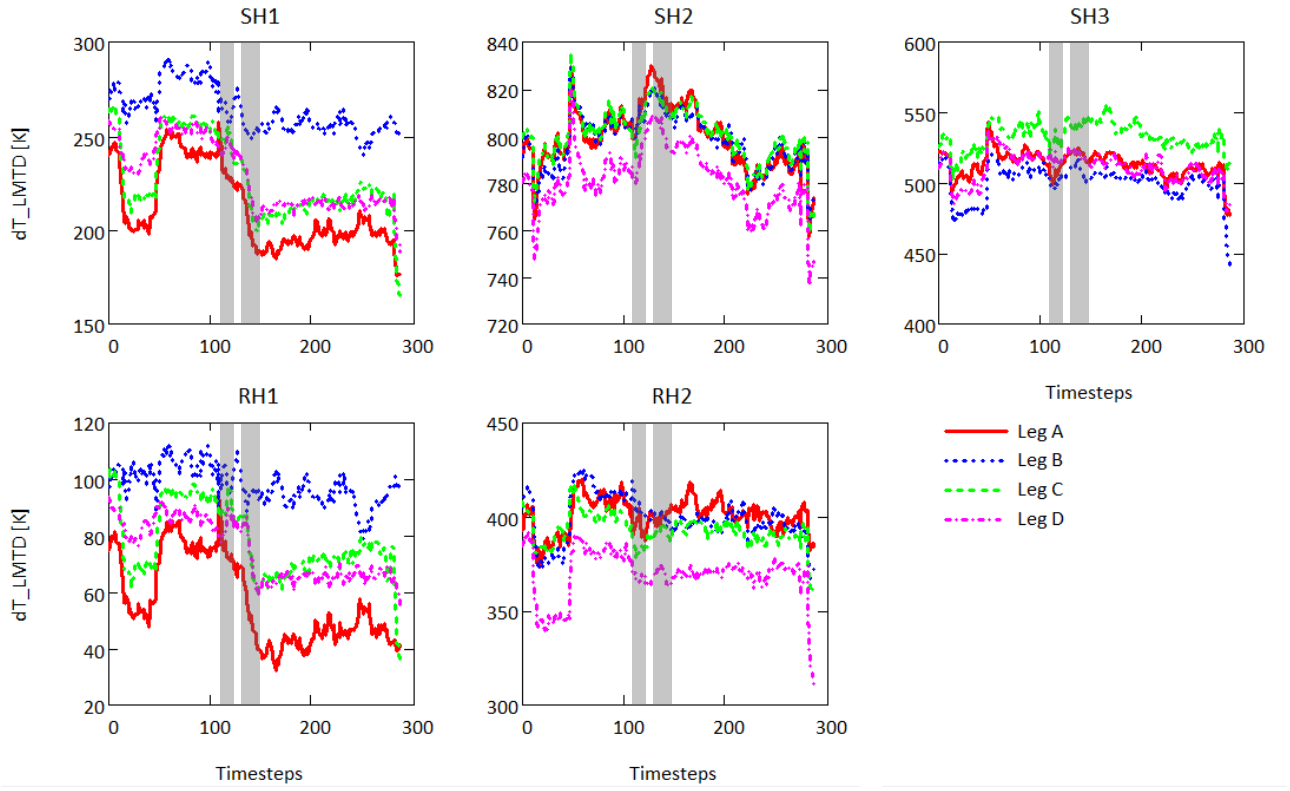


Figure 5-11: Log mean temperature difference across each leg

5.2 Comparison of soft sensing and data-driven approaches

The data-driven method evaluates the measurement sample of each timestep using a fault matrix to diagnose the condition of some unmeasured variables. The average computation time for each evaluation is about 0.05 milliseconds, being in the order of 10^7 times faster than the soft sensor. The model used to train the fault matrix is essentially the same as that used for the soft sensor approach and the results from these two methods are expected to exhibit the same trends. The data-driven results are therefore baselined against the soft sensor results throughout this section.

The faults included in the fault matrix are those representing changes in the conduction thermal resistance of each leg, as well as changes in all the process boundary conditions.

The diagnosis results for the conduction thermal resistances are shown in Figure 5-12.

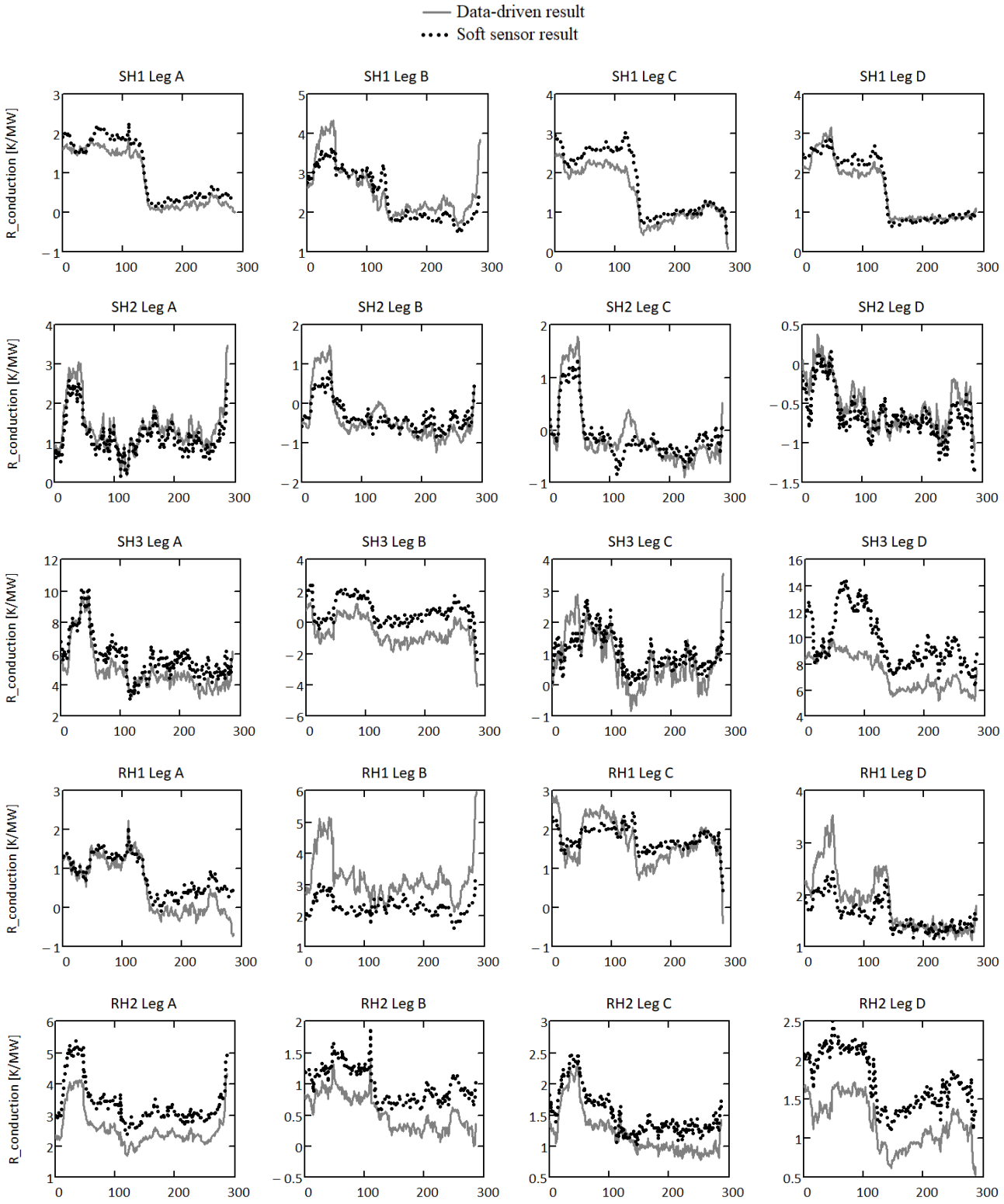


Figure 5-12: Conduction thermal resistance results from soft sensor and data-driven methods

The diagnosis results for the unmeasured process boundary condition of flue gas inlet temperature are shown in Figure 5-13.

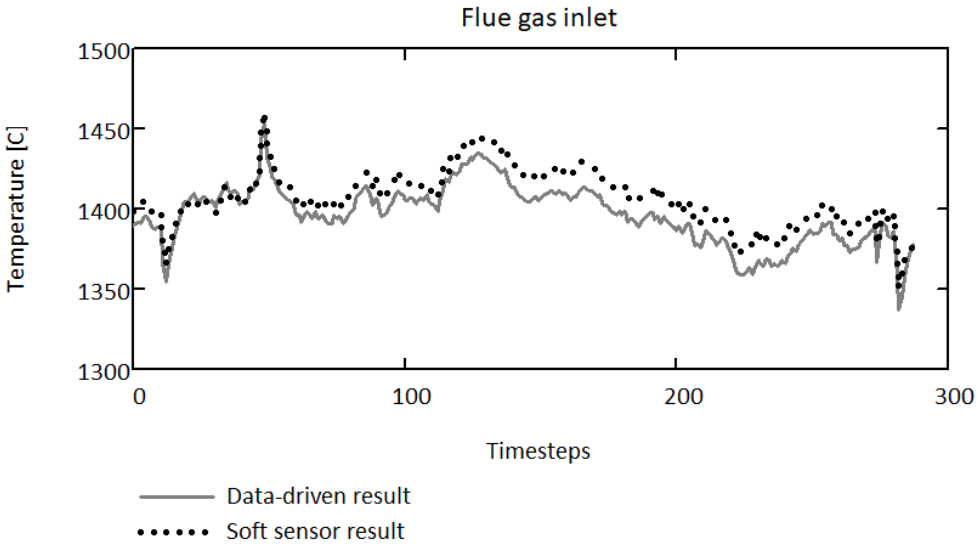


Figure 5-13: Flue gas inlet temperature results from soft sensor and data-driven method

Relative changes in the conduction thermal resistance and flue gas inlet temperature results are closely correlated with the soft sensor results. There is a discrepancy between the absolute values of the results from the soft sensor and data-driven methods. This is a possible result of process nonlinearities as the data-driven method approximates the process interactions to be linear around the NOP. Nonlinearity can exist in the fault direction, as well as the rate of propagation along this direction relative to fault magnitude. The effect of these nonlinearities on the results of any one fault leads to a change in the diagnosis of other faults. This is because any measurement space movement not correctly attributed to a specified fault must be described by those remaining in the fault matrix.

The discrepancy between the soft sensor and data-driven results of the conduction thermal resistance of some legs are accentuated during load changes. This is possibly because the effect of nonlinearities is more prominent for larger movements in the measurement space. A possible solution to this would be to move the NOP closer to the values expected at corresponding loads while still retaining the fault vectors of process boundary condition changes in the fault matrix. The offset between the NOP and the actual movement will then be captured in the process boundary condition fault indices while reducing the magnitude of the measurement deviation from the NOP.

The diagnosis results for the measured process boundary conditions are shown in Figure 5-14. It is of interest here to note that their fault vectors are however not only along the measurement dimension, thereby accounting for their influence on other measurements.

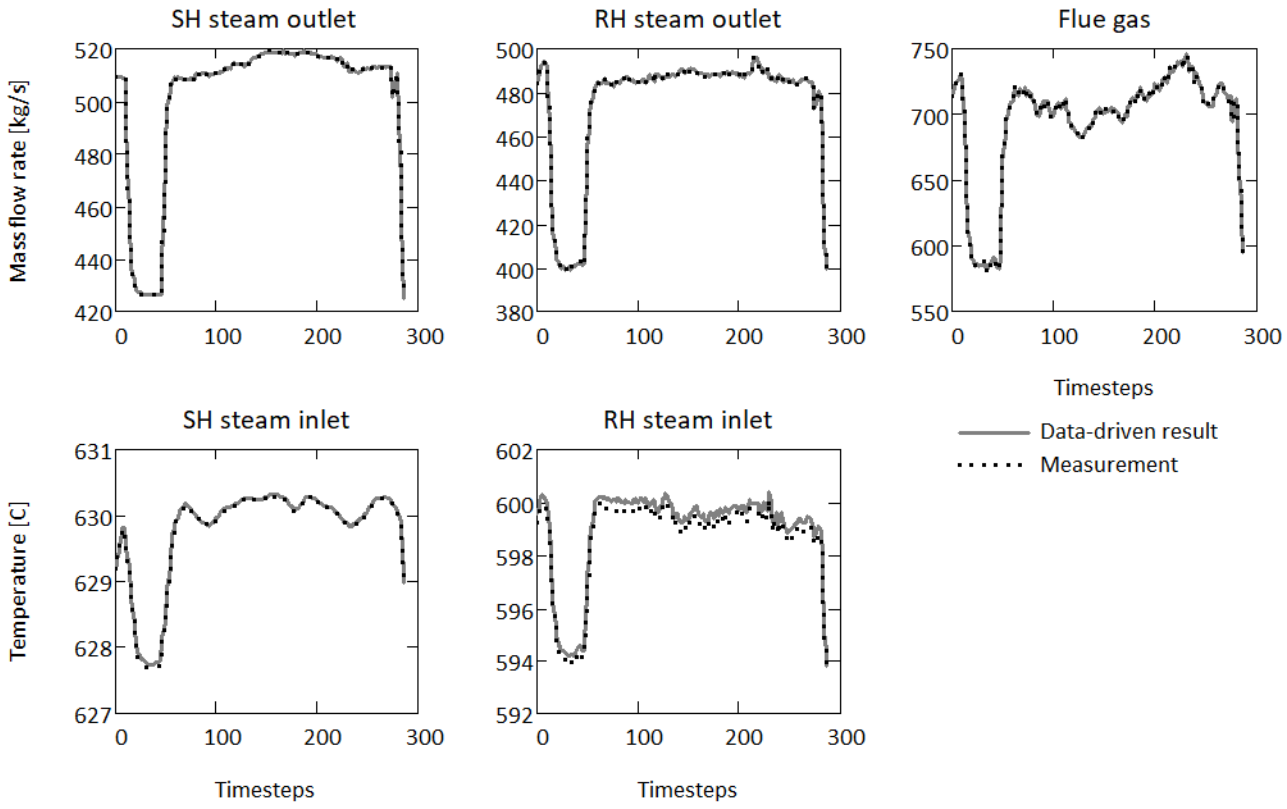


Figure 5-14: Measured process boundary conditions compared to data-driven results

The measured process boundary conditions are accurately diagnosed by the data-driven method as each of these measurements only respond to changes in the corresponding process boundary condition.

Whereas the soft sensor made use of all 53 measurements to fully determine the model, the data-driven method can diagnose the conduction thermal resistance of individual legs using only 20 measurements. Including changes in process boundary conditions as faults, a linearly independent fault matrix is achieved with only 26 measurements. This is because, from the data-driven perspective, 12 of the measurements do not respond to any of these faults, 12 are redundant and 4 respond similarly only to one fault.

The final outlet steam temperature of both the superheat and reheat sections are controlled at a setpoint by changing spray water flow rates and therefore do not respond to any fault. The second

stage attemperator spray water maintains a fixed steam temperature difference between its inlet and outlet. Within the operational conditions simulated during training, negligible changes in the spray water flow rate is required to achieve this, even in the presence of faults. These are the 12 measurements that don't respond to any faults.

Within the conditions simulated during training, any two of the three measurements around a first or third stage attemperator can be used to obtain the third. This is because a mass and energy balance around an attemperator subsystem with known steam flow rate and spray water properties can be solved with only two of these measurements. The steam temperature across a second stage attemperator is controlled at a fixed offset. Either of these temperatures are therefore given by the other and this offset. These are the 12 redundant measurements.

For the conditions simulated during training, the four steam mass flow rate measurements at the final superheater outlet only respond significantly to changes in the total steam mass flow rate. The response of the four measurements are similar, each measuring approximately a quarter of the total steam mass flow rate. Three of these measurements are therefore also redundant as any one of these measurements can be used to obtain the remaining three. Instead of discarding three of these measurements, all four are lumped to represent a single measurement of the total flow.

It should be noted that, although some measurements behave redundantly for the conditions simulated during training, severe flow maldistributions or a wide spectrum of operational ranges can cause these measurements to become non-redundant.

Although the data-driven method can provide diagnosis with less measurements required to fully determine the soft sensor model, the same can actually also be achieved with the soft sensor. The data-driven method can accurately diagnose the conduction resistances of individual legs using only interstage steam temperature and spray water flow measurements with fixed process boundary conditions and therefore a fixed NOP. This is essentially the same as fixing the process boundary conditions in the soft sensor model at their normal values and simulating the model to calculate the conduction resistances. When the process boundary conditions change, the NOP will move, and an error will be introduced in the data-driven conduction resistance results according to the influence of the process boundary conditions on the interstage steam and spray water measurements. This is because the measurement space movement must be described by changes in conduction resistance only when the fault matrix is only composed of these fault vectors. This is alike to simulating the soft sensor model without updating the process boundary conditions. Introducing changes in process boundary conditions as faults in the fault matrix is alike to updating the process boundary conditions in the soft sensor model.

The linear dependency that arises when attempting to diagnose discretized flue gas inlet temperature faults is alike to the soft sensor model having more unknowns than equations when attempting to calculate the discretized flue gas inlet temperatures. The availability of discretized flue gas temperature measurements at the economiser inlet provides the unknowns required to enable the soft sensor model to provide a unique solution. These measurements should therefore also provide the dimensions required to obtain a linearly independent fault matrix. To test this, the data-driven training model was modified to output the discretized flue gas exit temperature to postulate the effect of including these variables as measurements. Changes were then imposed on the discretized flue gas inlet temperatures. The linear dependency of the resulting fault matrix is then evaluated using a reciprocal analysis with a disturbance angle of five degrees. The resulting index matrix is shown in Figure 5-15.

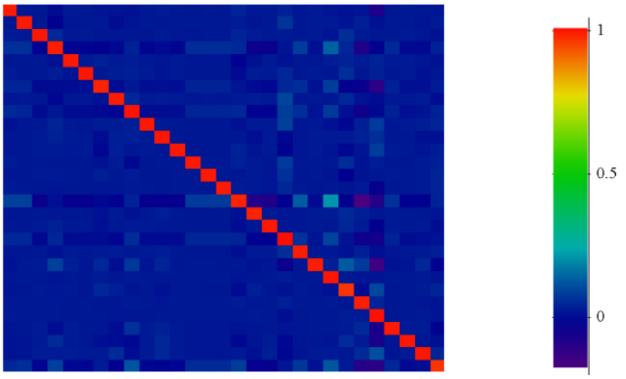


Figure 5-15: Reciprocal analysis for discretized flue gas inlet faults with discretized flue gas exit measurements

By comparing Figure 4-13, which was generated with only a one degree disturbance angle, with Figure 5-15 it is evident that the introduction of discretized flue gas exit temperature measurements provides the dimensions necessary to diagnose discretized flue gas inlet temperature changes. This will provide an estimate of the temperature profile into the convective pass, enabling the effects of online process operations on this profile to be monitored.

The reciprocal analysis method used in this study provides a practical way of quantifying the extent of linear dependency. Analysis of the linear dependency of fault vectors can be used to establish the achievable monitoring resolution for a given measurement configuration.

The fault matrix describes the reaction of the system to faults and is therefore robust to minor modelling uncertainties. Supposing that the model represented the actual system, it has been shown that the data-driven method can accurately diagnose system operation without the use of a model if training data is available.

The data-driven method training is based on a model that assumes perfect steady state control. When the process variables deviate from an in-control state, the steady state model used for training no longer represents the actual process and the data-driven method can become disordered. The diagnosis results can also respond erroneously when controls act on the system that the data-driven method was not exposed to during its training, especially when these controls cause abrupt turning points or non-smooth responses in the process variables. Examples of this are spray water flows being limited to maintain a minimum degree of superheat, or saturation of spray water valve flow capacity. The soft sensor does not incorporate process controls in the model and is insensitive to out of control operation. The measurements found to be redundant for the data-driven method and therefore excluded from its inputs are not redundant for the soft sensor. It is therefore expected that the data-driven and soft sensor results should differ as these methods do not consider the same measurements. Also, these results are essentially based on different models as the data-driven model incorporates process controls.

While the data-driven method allows the location and severity of ash fouling to be monitored, it also provides monitoring outcomes related to some damage mechanisms. The flue gas inlet temperature can be related to overheating damage. Tracking the post soot blowing conduction thermal resistance results might provide information about the average oxide layer behaviour of each leg.

The monitoring outcomes of the soft sensor provide more detailed information about the flow and temperature of boiler fluids throughout the convective pass. This enables a more accurate estimation of tube wall temperatures across each leg which is related to overheating damage. This information can also be related to the average erosion rates expected across the individual legs.

Other than providing information related to boiler material health, both monitoring frameworks enable the effectiveness of soot blowing to be analysed for individual legs. The ash build-up characteristics can be studied and the impact of different coal qualities, milling performance, firing configurations and operating actions on fouling tendencies can be quantified. The monitoring outcomes also serve as an additional tool for root cause analysis of boiler problems such as thermal index excursions on headers, attemperator over spraying, ID fan saturation or boiler tube failures.

The soft sensor results are more accurate as it accounts for process nonlinearities. Although the data-driven method is less accurate, it is also able to serve as a soft sensor because it can estimate the absolute values of the unmeasured variables. The relative changes in the results from the two methods correlate very well. Although the absolute values of boiler variables are beneficial, the relative changes in these variables are equally useful from an online condition monitoring perspective (Kim et al. 2014).

The data-driven method has much lower computational requirements and does not need the continuous simulation of specialised software as is the case with the soft sensor method. The specialised software is however required during the training phase if labelled historical operating data is not available. The simplified training methodology allowed by this method however requires a minimal amount of simulations. Once the training is complete, the data-driven method can be easily implemented on a wide variety of platforms, relieving software license constraints.

By monitoring unmeasured variables related to damage mechanisms online, the effect of operating actions on boiler health and performance can be observed, events that trigger accelerated damage can be identified and the progression of boiler condition can be predicted.

6. Conclusion

The results presented in this study show that standard process measurements installed on modern boilers enable unmeasured fluid and heat exchanger variables related to performance and damage mechanisms to be monitored online. This is achieved by using a model built on simplified representations of boiler thermofluid process knowledge. The process model is used both directly as a soft sensor and indirectly to train a data-driven monitoring method.

The soft sensor provides the unmeasured variables through the solution to the model equations given the measured variables as inputs. The soft sensor results include the average fluid variables through and surrounding each leg. It further provides the contributions of external convection, radiation and conduction to the overall thermal resistance of each heat exchanger leg.

A data-driven method is developed in this study to diagnose multiple simultaneous faults and is trained by imposing a change in selected variables on the process model. Only one training example is used for each fault at a given load. The data-driven method establishes the contribution of potential faults to the deviation of a measurement sample from the normal operating point. The diagnosis results include the conduction thermal resistance of each leg and the flue gas inlet temperature.

The soft sensor and data-driven methods are equivalent in the measurements required to be fully defined. The detailed results provided by the soft sensor are more relevant for condition prognosis than the data-driven method. The soft sensor results are also more accurate as it accounts for nonlinearities in variable interactions. The data-driven method is therefore inferior in terms of the variety and accuracy of its results. The relative changes in the joint results from the two methods however correlate very well. Although knowledge of the absolute values of boiler variables are beneficial, the relative changes in these variables are equally useful from an online monitoring perspective. The data-driven method has much lower computational requirements and does not need the continuous simulation of specialised software. This provides a practical online monitoring tool that is easily implementable on a variety of software platforms.

The practical implementation of these monitoring frameworks allows the interaction of key unmeasured variables in boilers to be observed in real time, providing knowledge that is useful for optimizing boiler operation and maintenance.

7. Recommendations

Dealing with non-smooth process interactions is a potential area for improvement of the data-driven method. Such interactions can be caused by conditional process controls and flow capacity limitations. Developing different fault matrices for different regions in the measurement space might provide a possible solution to this.

It was found that minimal flow occurs in connecting pipes and ducts and it is therefore recommended that unless this information provides valuable prognostic insight, momentum conservation need not be considered in the models to simplify computational and software requirements. However, this might change in cases of severe maldistribution between connected legs.

Measurement space movement necessarily occurs because some variables are functions of others and the movement directions are therefore defined by these functions. Incorporating information about the derivatives of these functions in a data-driven method might therefore provide a solution for dealing with nonlinear measurement space movement.

The soft sensor method can be extended to evaluate the average effects of changing process conditions on tube metal temperatures. By establishing the average internal oxide layer thickness during outage inspections, the conduction resistance being monitored online can be used to estimate the average fouling thickness. This information will provide better insight for online calculation of tube metal temperatures of individual heat exchanger legs.

The data-driven method developed in this study provides a simplified linear model of process behaviour in response to variable changes in the form of a matrix. The variable changes included in the development of the matrix do not necessarily need to correspond to faults and can be used to monitor changes in any variable of interest. This method can find applicability in other thermofluid and non-thermofluid processes of which variable interactions are represented by continuous functions. Because the model is stored in a matrix, it can easily be incorporated into the process control software for automated fault diagnosis. This can assist operators with conducting a first-line investigation before a system expert is called out to assist with anomalous process behaviour. The method can prove particularly useful in emergency situations where quicker diagnosis is required than that achievable using the available process models.

8. List of References

Alcala, C.F. & Qin, S.J., 2011. Analysis and generalization of fault diagnosis methods for process monitoring. *Journal of Process Control*, 21(3), pp.322–330. Available at: http://dx.doi.org/10.1016/j.jprocont.2010.10.005.

Alcala, C.F. & Qin, S.J., 2009. Reconstruction-based contribution for process monitoring. *Automatica*, 45(7), pp.1593–1600. Available at: http://dx.doi.org/10.1016/j.automatica.2009.02.027.

Alnaimi, F.B.I. & Al-Kayiem, H.H., 2011. Artificial intelligent system for steam boiler diagnosis based on superheater monitoring. *Journal of Applied Sciences*, 11(9), pp.1566–1572.

ASME, 2008. ASME PTC4. *Asme*, 2008.

Cantrell, C. & Idem, S., 2010. On-Line Performance Model of the Convection Passes of a Pulverized Coal Boiler. *Heat Transfer Engineering*, 31(14), pp.1173–1183. Available at: http://www.tandfonline.com/doi/abs/10.1080/01457631003689328.

Cardoso, B.R. et al., 2012. Microstructural degradation of boiler tubes due to the presence of internal oxide layer. *Journal of Materials Research and Technology*, 1(2), pp.109–116. Available at: http://dx.doi.org/10.1016/S2238-7854(12)70020-0.

Chiang, L.H. et al., 2015. Diagnosis of multiple and unknown faults using the causal map and multivariate statistics. *Journal of Process Control*, 28, pp.27–39. Available at: http://dx.doi.org/10.1016/j.jprocont.2015.02.004.

Clark, R.N.C.R.N., 1979. The dedicated observer approach to instrument failure detection. *1979 18th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, 18, pp.237–241.

Colebrook, C.F., 1939. Turbulent flow in pipes, with particular reference to the transition region between the smooth and rough pipe laws. *Journal of the Institution of Civil engineers*, 12(8), pp.393–422.

Coleman, K., 2011a. Boiler and Heat Recovery Steam Generator Tube Failures - Theory and Practice: Volume 1: Fundamentals, Volume 2: Water-Touched Tubes, and Volume 3: Steam-Touched Tubes - Product Abstract. *Epri*, 1(1023063), pp.2–3.

Coleman, K., 2011b. Boiler and Heat Recovery Steam Generator Tube Failures - Theory and Practice: Volume 1: Fundamentals, Volume 2: Water-Touched Tubes, and Volume 3: Steam-Touched Tubes - Product Abstract. *Epri*, 2(1023063), pp.2–3.

Cybenko, G., 1989. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems*.

Das, S.K. et al., 2006. Analytical model for erosion behaviour of impacted fly-ash particles on coal-fired boiler components. *Sadhana - Academy Proceedings in Engineering Sciences*, 31(5), pp.583–595.

Díez, L.I., Cortés, C. & Campo, A., 2005. Modelling of pulverized coal boilers: Review and validation of on-line simulation techniques. *Applied Thermal Engineering*, 25(10), pp.1516–1533.

Ding, S.X., 2008. *Model-based Fault Diagnosis Techniques - Design Schemes, Algorithms and Tools*,

Dooley, B. & Chang, P.S., 2000. The current state of boiler tube failures in fossil plants. *Power Plant Chemistry*, 2(4), pp.197–203.

Dunia, R. & Qin, S.J., 1998. Joint diagnosis of process and sensor faults using principal component analysis. *Control Engineering Practice*, 6(4), pp.457–469.

Dunia, R. & Qin, S.J., 1998. Subspace approach to multidimensional fault identification and

reconstruction. *AIChE Journal*, 44(8), pp.1813–1831.

Eskom, 2017. *Eskom Integrated Report*,

Fathi, Z., Ramirez, W.F. & Korbicz, J., 1993. Analytical and knowledge-based redundancy for fault diagnosis in process plants. *AIChE Journal*, 39(1), pp.42–56.

Fortuna, L. et al., 2007. *Soft Sensors for Monitoring and Control of Industrial Processes*,

Frank, P.M., 1990. Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge -Based Redundancy, A Survey and Some New Results. *Automatica*, 26(3), pp.459–474.

Frank, P.M., Ding, S.X. & Marcu, T., 2000. Model Based Fault Diagnosis in Technical Processes. *Trans. Inst. MC / Millennium issue*, 22 (1), pp.57–101.

Gertler, J., 1991. Analytical Redundancy Methods in Fault Detection and Isolation - Survey and Synthesis. *IFAC Proceedings Volumes*, 24(6), pp.9–21. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1474667017511192.

Gertler, J.J., 1997. Fault Detection and Isolation using Parity Relations. *Control Engineering Practice*, 5(5), pp.653–661.

Gertler, J.J. & Monajemy, R., 1995. Generating directional residuals with dynamic parity relations. *Automatica*, 31(4), pp.627–635.

Gertler, J.J. & Singer, D., 1990. A New Structural Framework for Parity Equation-Based Failure Detection and Isolation. *Automatica*, 26, pp.381–388.

Govindsamy, R., 2014. *Thermal Performance Evaluation of Heat Exchangers in Pulverised Coal Boilers*. University of the Witwatersrand.

Grimson, E., 1937. Correlation and Utilization of New Data on Flow Resistance and Heat Transfer Cross-flow of Gases Over Tube Banks. *Trans. ASME*, 59, pp.583–594.

Haves, P., Salsbury, T.I. & Wright, J.A., 1996. Condition monitoring in HVAC subsystems using first principles models. *ASHRAE Transactions*, 102(1), pp.519–527.

He, T. et al., 2006. Process fault detection and diagnosis based on principal component analysis. *Fifth International Conference on Machine Learning and Cybernetics*, 1(August), pp.3551–3556.

Hilpert, R., 1933. Wärmeabgabe von geheizten Drähten und Rohren im Luftstrom. *Forschung auf dem Gebiet des Ingenieurwesens A*, 4(5), pp.215–224. Available at: https://doi.org/10.1007/BF02719754.

Holman, J., 2010. Heat Transfer. *Mc Graw Hill*, p.758. Available at: https://books.google.se/books?isbn=0073529362.

Hoppenstedt, B. et al., 2018. Techniques and Emerging Trends for State of the Art Equipment Maintenance Systems—A Bibliometric Analysis. *Applied Sciences*, 8, pp.1–29. Available at: http://www.mdpi.com/2076-3417/8/6/918.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), pp.417–441.

Incropera, F.P. et al., 2007. *Fundamentals of Heat and Mass Transfer*, Available at: http://www.google.com/patents?hl=en&lr=&vid=USPAT5328671&id=rb8lAAAAEBAJ&oi=fnd&dq=Heat+and+Mass+Transfer&printsec=abstract%5Cnhttp://www.google.com/patents?hl=en&lr=&vid=USPAT5328671&id=rb8lAAAAEBAJ&oi=fnd&dq=Heat+and+mass+transfer&printsec=abstract%5Cnh.

International Energy Agency, 2016. Key world energy statistics. *Statistics*, p.80. Available at: http://www.iea.org/statistics/statisticssearch/.

Isermann, R., 2011. *Fault-Diagnosis Applications*, Darmstadt: Springer.

Isermann, R., 1984. Process fault detection based on modeling and estimation methods-A survey.

Automatica, 20(4), pp.387–404.

Ismail, F.B. et al., 2016. Early tube leak detection system for steam boiler at KEV power plant. , 00006, pp.4–8.

Jämsä-Jounela, S.L., 2007. Future trends in process automation. *Annual Reviews in Control*, 31(2), pp.211–220.

Kambhatla, N. & Leen, T.K., 1997. Dimension Reduction by Local Principal Component Analysis. *Neural Computation*, 9(7), pp.1493–1516. Available at: http://www.mitpressjournals.org/doi/abs/10.1162/neco.1997.9.7.1493.

Katipamula, S. & Brambley, M., 2005. Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part I. *HVAC&R Research*, 11(1), pp.3–25. Available at: http://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391123.

Kays, W.M. & London, A.L., 1984. *Compact Heat Exchangers*,

Kim, H., Na, M.G. & Heo, G., 2014. Application of monitoring, diagnosis, and prognosis in thermal performance analysis for nuclear power plants. *Nuclear Engineering and Technology*, 46(6), pp.737–752. Available at: http://dx.doi.org/10.5516/NET.04.2014.720.

Li, G., Qin, S.J. & Chai, T., 2014. Multi-directional reconstruction based contributions for root-cause diagnosis of dynamic processes. *Proceedings of the American Control Conference*, pp.3500–3505.

Li, H., Yu, D. & Braun, J.E., 2011. A Review of Virtual Sensing Technology and Application in Building Systems. *HVAC&R Research*, 9669(November 2014), pp.37–41.

Ma, L. et al., 2014. ANN-Based Diagnosis of Boiler Four-tube Leakage Faults Under Different Loads and Operating Modes. , (3), pp.88–92.

MacGregor, J.F. et al., 1994. Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, 40(5), pp.826–838.

MacGregor, J.F. & Kourti, T., 1995. Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), pp.403–414.

Mbabazi, J.G., Sheer, T.J. & Shandu, R., 2004. A model to predict erosion on mild steel surfaces impacted by boiler fly ash particles. *Wear*, 257(5–6), pp.612–624.

Miller, P., Swanson, R.E. & Heckler, C.E., 1998. Contribution plots: a missing link in multivariate quality control. *Journal of Applied Mathematics and Computer Science*.

Nielson, F.S., Danesi, P. and Radhakrishnan, M.., 2012. Modern Boiler Design. *BWE Energy India*, (January). Available at: http://www.bwe.dk/download/articles_pdf/modernboilerdesign2012.pdf.

Nomikos, P. & MacGregor, J.F., 1995. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, 37(1), pp.41–59.

Pattanayak, L. et al., 2015. Optimization of sootblowing frequency to improve boiler performance and reduce combustion pollution. *Clean Technologies and Environmental Policy*, pp.1897–1906. Available at: http://dx.doi.org/10.1007/s10098-015-0906-0.

Pearson, K., 1901. LIII. *On lines and planes of closest fit to systems of points in space*. *Philosophical Magazine Series 6*, 2(11), pp.559–572. Available at: http://www.tandfonline.com/doi/abs/10.1080/14786440109462720.

Pena, B., Teruel, E. & Diez, L.I., 2013. Towards soot-blowing optimization in superheaters. *Applied Thermal Engineering,* 61, pp.737-746.

Qin, S.J., 2003. Statistical process monitoring : basics and beyond. , pp.480–502.

Raich, A. & Cinar, A., 1996. Statistical Process Monitoring and Disturbance Diagnosis in Multivariate Continuous Processes.Pdf. , 42(4). Available at:

http://xlink.rsc.org/?DOI=C3AY41907J.

Rao, C.R., 1967. Calculus of Generalized Inverses of Matrices Part I: General Theory. *Sankhya*, 29(3), pp.317–342. Available at: http://www.jstor.org/stable/25049483.

Rato, T.J. & Reis, M.S., 2014. Non-causal data-driven monitoring of the process correlation structure: A comparison study with new methods. *Computers & Chemical Engineering*, 71, pp.307–322. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0098135414002555.

Rennels, D.C. & Hudson, H.M., 2012. *Pipe Flow: A Practical and Comprehensive Guide*,

Robinson, D.J.S., 2011. Linear algebra. *World*, (Book, Whole). Available at: http://tue.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwdV3dS8MwED-cA1EUnB-1frFXH1q6dktyjzo3htMnRedTWNdUBmNM17F_37usrZsw6Ms1oSXhvpP7HUAU-oH3TycoacjXGIkg5pBIJaFKkxEaMr5pLEK0dcP95v2T_Oi0PtcQhpZLvziftPfgWXUXbaW40OvrD6SgjIkUp4v4wJJFuQIVZQW2jYMy9xKQV0KOA7f.

Rostek, K., Morytko, L. & Jankowska, A., 2015. Early detection and prediction of leaks in fluidized-bed boilers using artificial neural networks. *Energy*, 89, pp.914–923.

Seongkyu Yoon & MacGregor, J.F., 2001. Fault diagnosis with multivariate statistical models part I: using steady state fault signatures. *Journal of Process Control*, 11(4), pp.387–400.

Shi, Y., Wang, J. & Liu, Z., 2015. On-line monitoring of ash fouling and soot-blowing optimization for convective heat exchanger in coal- fi red power plant boiler. *Applied Thermal Engineering*, 78, pp.39–50. Available at: http://dx.doi.org/10.1016/j.applthermaleng.2014.12.002.

Si, F. et al., 2009. Inferential sensor for on-line monitoring of ammonium bisulfate formation temperature in coal-fired power plants. *Fuel Processing Technology*, 90(1), pp.56–66. Available at: http://dx.doi.org/10.1016/j.fuproc.2008.07.015.

Taler, J., Trojan, M. & Taler, D., 2009. Assessment of ash fouling and slagging in coal fired utility boilers. *International Conference on Heat Exchanger Fouling and Cleaning VIII*, 2009, pp.103–112.

Tidriri, K. et al., 2016. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. *Annual Reviews in Control*, 42, pp.63–81. Available at: http://dx.doi.org/10.1016/j.arcontrol.2016.09.008.

Trefethen, L.N. & Bau III, D., 1997. Numerical linear algebra. *Numerical Linear Algebra with Applications*.

Trojan, M. & Taler, D., 2015. Thermal simulation of superheaters taking into account the processes occurring on the side of the steam and flue gas. *Fuel*, 150, pp.75–87. Available at: http://dx.doi.org/10.1016/j.fuel.2015.01.095.

Valero, A. & Cortbs, C., 1996. Ash Fouling in Coal-Fired Utility Boilers. Monitoring and Optimization of On-Load Cleaning. , 22(96), pp.189–200.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., et al., 2003. A review of process fault detection and diagnosis. *Computers & Chemical Engineering*, 27(3), pp.293–311.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., et al., 2003. A review of process fault detection and diagnosis Part III: Process history based methods. *Computers & Chemical Engineering*, 27(3), pp.293–311.

Venkatasubramanian, V., Rengaswamy, R. & Kavuri, S.N., 2003. A review of process fault detection and diagnosis part II: Qualitative models and search strategies. *Computers and Chemical Engineering*, 27(3), pp.313–326.

Weisbach, J.L., 1845. *Lehrbuch der Ingenieur-und Maschinen-Mechanik: Theoretische Mechanik*, Druck und Verlag von Friedrich Vieweg und Sohn.

Xu, L., Khan, J.A. & Chen, Z., 2000. Thermal load deviation model for superheater and reheater of a utility boiler. *Applied Thermal Engineering*, 20(6), pp.545–558.

Yin, S. et al., 2012. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control*, 22(9), pp.1567–1581. Available at: http://dx.doi.org/10.1016/j.jprocont.2012.06.009.

Yin, S. et al., 2014. A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11), pp.6414–6428.

Yu, J., Jang, J., Yoo, J., Park, J.H., et al., 2016. A clustering-based fault detection method for steam boiler tube in thermal power plant. *Journal of Electrical Engineering and Technology*, 11(4), pp.848–859.

Yu, J., Jang, J., Yoo, J. & Kim, S., 2016. Fault Detection Method for Steam Boiler Tube Using Mahalanobis Distance. , V(3), pp.246–252.

Yu, J. et al., 2015. Leakage Detection of Steam Boiler Tube in Thermal Power Plant Using Principal Component Analysis. , (2008), pp.1–8.

Yue, H.H. & Qin, S.J., 2001. Reconstruction-Based Fault Identification Using a Combined Index. *Industrial & Engineering Chemistry Research*, 40(20), pp.4403–4414. Available at: http://pubs.acs.org/doi/abs/10.1021/ie000141+.

Zhou, Z. et al., 2016. Analysis of PCA-based reconstruction method for fault diagnosis. *Industrial & Engineering Chemistry Research*, p.acs.iecr.5b04822. Available at: http://pubs.acs.org/doi/abs/10.1021/acs.iecr.5b04822.

Appendix A. Flue gas property curves

The flue gas properties determined as described in Section 3.1.1 is shown here. The flue gas enthalpy-temperature relation derived for the typical full load composition is compared to that of air in Figure 8-1.

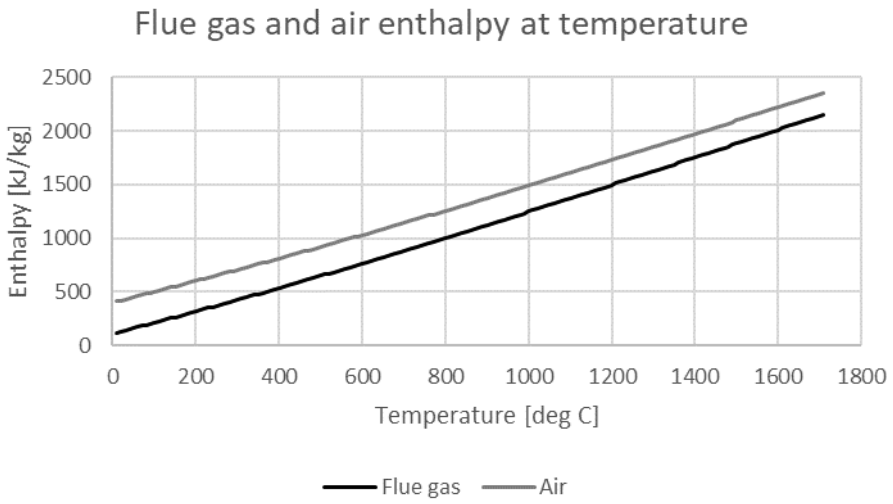


Figure 8-1: Flue gas enthalpy temperature relation compared to that of air

The difference between the flue gas enthalpy-temperature relation at full load and 70% load is shown in Figure 8-2 as a percentage. This difference is brought about by the difference in flue gas composition at lower loads because of different air fuel ratios and combustion characteristics.

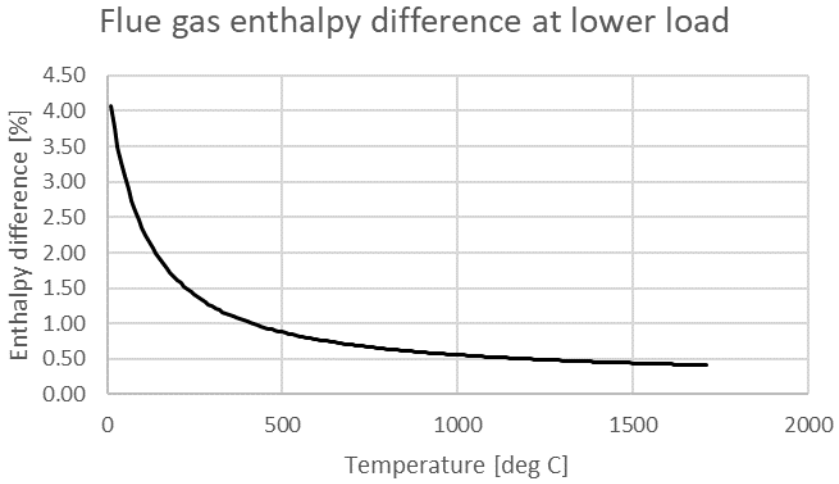


Figure 8-2: Difference in flue gas enthalpy at temperature for different loads

The flue gas density-temperature relation obtained for the typical full load composition is compared to that of air in Figure 8-3.

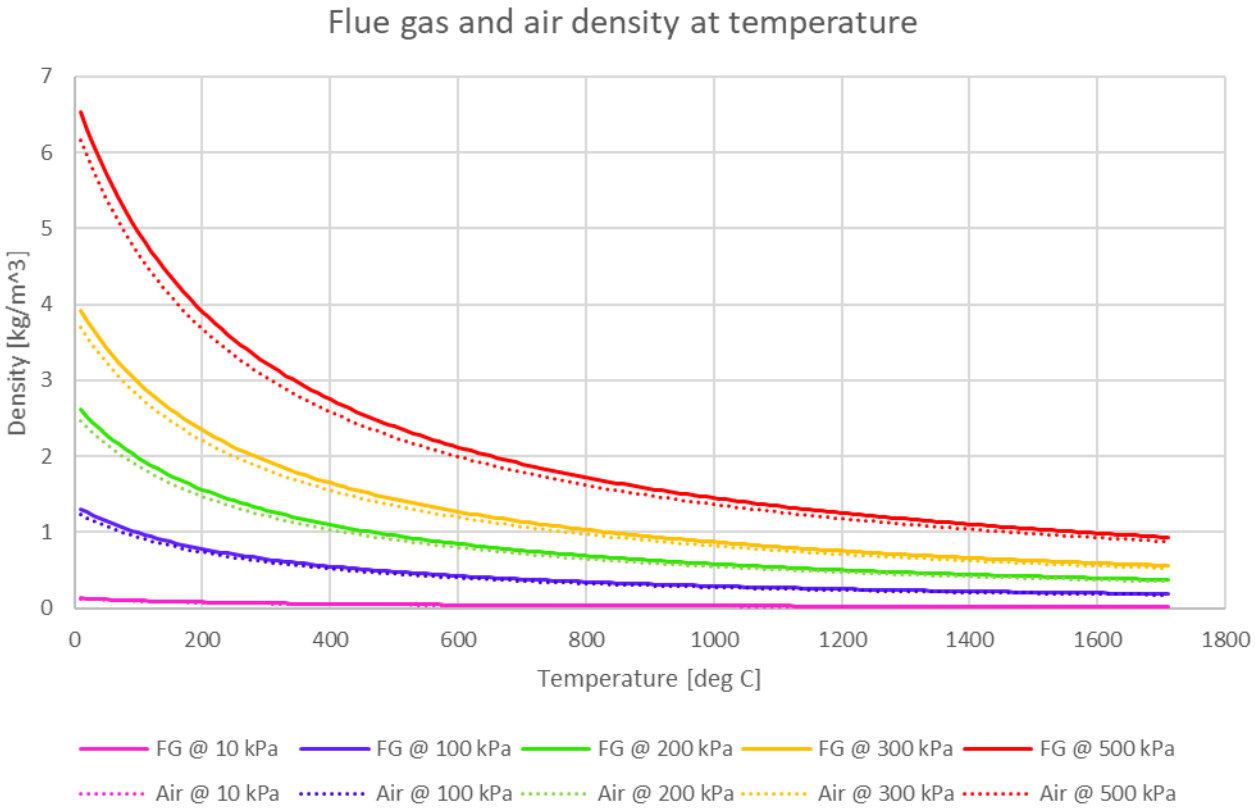


Figure 8-3: Flue gas and air density as a function of temperature for various pressures

Appendix B. Soft sensor for total flue gas mass flow rate

Either of the following approaches can be used to soft sense the total flue gas mass flow rate.

A mass balance of flows entering and leaving the furnace area yields the total flue gas mass flow rate. Although coal and air flows into the boiler are measured, the fraction of ash too heavy to be entrained in the flue gas flowing to the convective pass is not. The bottom ash flow can change with operational factors like milling performance and coal composition. A continuous measurement of coal composition is not available. A fraction of 10% of the total coal ash content is suggested for South African CFPPs (Govindsamy 2014).

Since the flue gas temperature is measured both up- and downstream of the economiser, and the heat transfer rate to the water in the economiser is known from other measurements, a mass and energy balance around the economiser provides another approach to determine the flue gas mass flow rate. Assuming no losses or directed radiation, it can be written as

$$\dot{m}_{fluegas} = \frac{\dot{m}_{water}\left(h_{water_{exit}} - h_{water_{inlet}}\right)}{h_{fluegas_{inlet}} - h_{fluegas_{exit}}} \tag{4.11}$$

The principals employed for the flue gas temperature measurement around the economiser are prone to errors.

For the condition monitoring approach employed in this research, accurately capturing relative changes in monitored variables take preference over accurate quantification of the absolute values of these variables. Flue gas heat losses, directed radiation and measurement errors should remain relatively constant during typical operation, while greater variation is expected in the bottom ash flow rate. The mass and energy balance around the economiser is therefore chosen to obtain the flue gas mass flow rate in this study.

Appendix C. Appearance of soft sensor model in modelling software

The steam and flue gas side models as it appears in the modelling software is shown in Figure 8-4 to Figure 8-6. The model is split into three figures to allow for reasonable resolution.

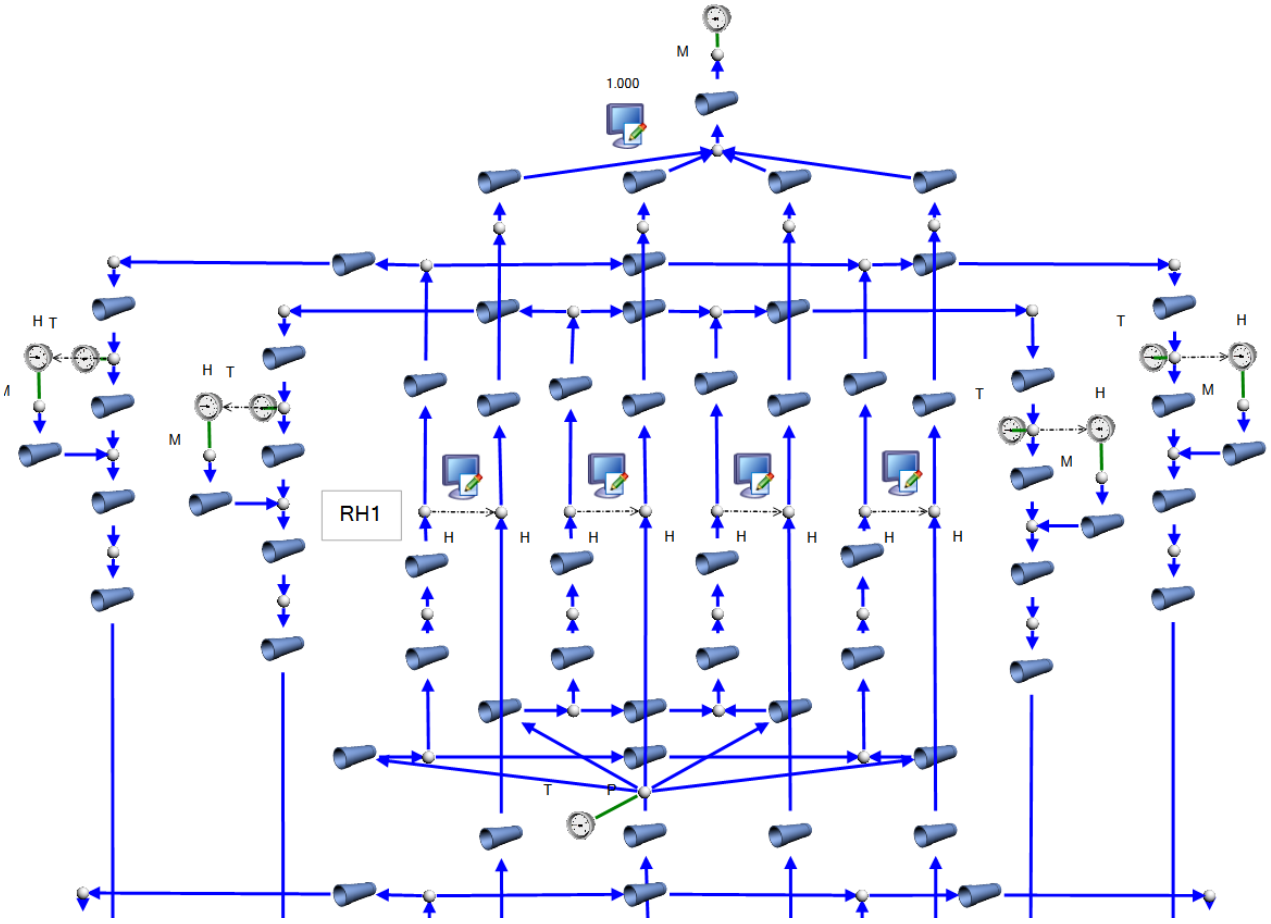


Figure 8-4: Steam and flue gas soft sensor model around RH1

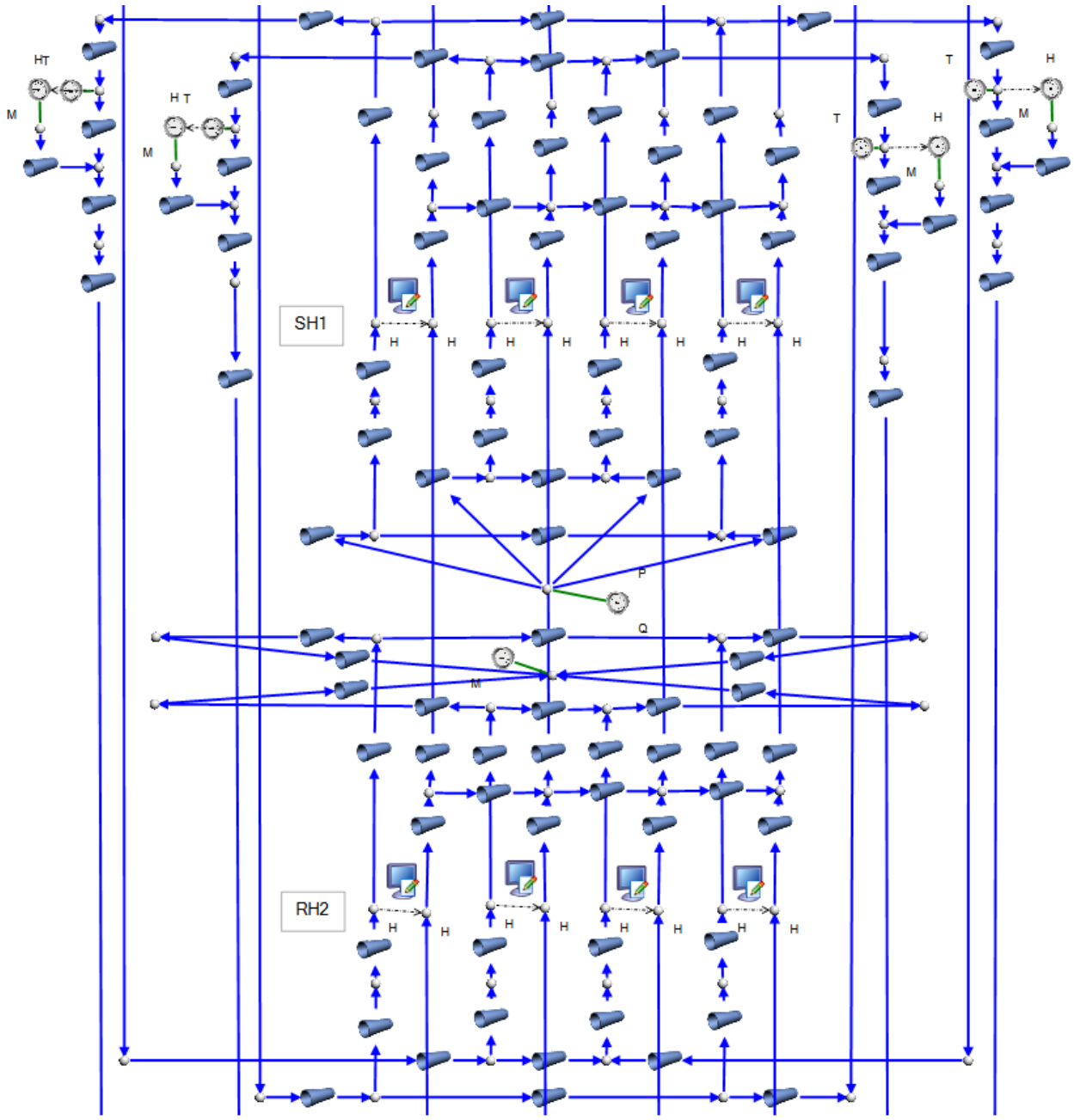


Figure 8-5: Steam and flue gas soft sensor model around RH2 and SH1

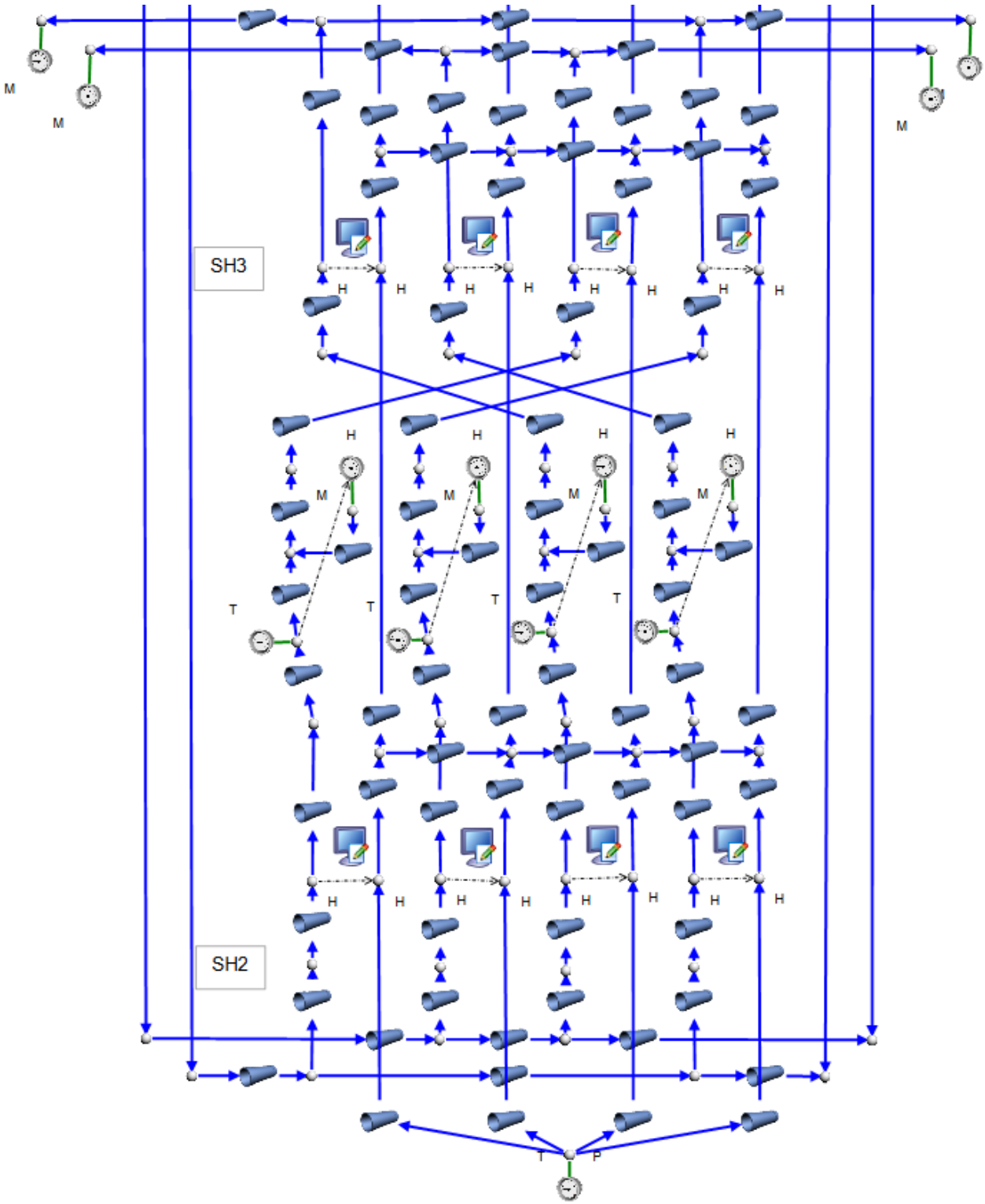


Figure 8-6: Steam and flue gas soft sensor model around SH2 and SH3

Appendix D. Verification of convection and radiation resistance scaling in modelling software

The approach for scaling the convection and radiation thermal resistance with changing process conditions is implemented in the modelling software using a custom component developed by Associate Professor Wim Fuls at the University of Cape Town. The working of this component is therefore verified with a hand calculation. The hand calculation working is also included later in this appendix. The design values for an arbitrary heat exchanger is given in Table 8-1 along with three cases of operation at off design conditions used for the verification exercise. The overall heat transfer coefficient outputs of these models are then compared.

Table 8-1: Design and operating values used for model verification

Parameter	Design value	Operating values		
		Case 1	Case 2	Case 3
Radiation heat transfer portion [%]	20	20	20	20
Steam flow rate [kg/s]	500	450	550	400
Steam pressure [MPa]	19	19	19	17
Steam inlet temperature [deg C]	400	390	410	420
Steam exit temperature [deg C]	440	423	456	458
Flue gas flow rate [kg/s]	800	750	850	900
Flue gas pressure [kPa]	100	100	100	80
Flue gas inlet temperature [deg C]	850	800	900	700

The difference in the overall heat transfer coefficient outputs from the model and hand calculation is shown in Table 8-2 for the design case and the three operating cases.

Table 8-2: Comparison of overall heat transfer coefficients from model output and hand calculations

Case	Overall heat transfer coefficient from model [W/K]	Overall heat transfer coefficient from hand calculation [W/K]	Difference [%]
Design	256950	256863	0.034
1	240396	240495	0.041
2	273261	273369	0.040
3	251943	252048	0.042

The difference, being within 0.05%, is acceptable and the custom component built in the modelling software is considered verified.

The hand calculation script for case 3 is shown below:

Inputs

Design conditions

Radiation heat transfer portion $Rr := 0.2$

Flue gas

Mass flow rate $m\_fg\_D := 800 \frac{kg}{s}$

Inlet temperature $T\_fg\_in := 850\ °C$

Exit temperature $T\_fg\_ex = 747.198 \cdot °C$

Pressure $p\_fg\_D := 100kPa$

Steam

Mass flow rate $m\_st\_D := 500 \frac{kg}{s}$

Inlet temperature $T\_st\_in := 400\ °C$

Exit temperature $T\_st\_ex := 440\ °C$

Pressure $p\_st\_D := 19MPa$

Operating conditions

Flue gas

Mass flow rate $m\_fg := 900 \frac{kg}{s}$

Inlet temperature $T\_fg\_in\_act := 700\ °C$

Exit temperature $T\_fg\_ex\_act = 644.416 \cdot °C$ (Obtained iteratively)

Pressure $p\_fg\_act := 80kPa$

Steam

Inlet temperature $T\_st\_in\_act := 420\ °C$

Exit temperature $T\_st\_ex\_act = 457.881 \cdot °C$ (Obtained iteratively)

Pressure $p\_st\_act := 17MPa$

Calculations

Design calculations

Steam

Average steam temperature $T\_st\_D := \frac{T\_st\_in + T\_st\_ex}{2} = 420 \cdot °C$

Inlet enthalpy $h\_st\_in := h_{steam}(p\_st\_D, T\_st\_in, "", "", "") = 2852.773 \cdot \frac{kJ}{kg}$

Exit enthalpy $h\_st\_ex := h_{steam}(p\_st\_D, T\_st\_ex, "", "", "") = 3042.466 \cdot \frac{kJ}{kg}$

Heat transfer rate $Q := m\_st\_D \cdot (h\_st\_ex - h\_st\_in) = 94.846 \cdot MW$

Specific heat $c\_st\_D := \frac{h\_st\_ex - h\_st\_in}{T\_st\_ex - T\_st\_in} = 4.742 \cdot \frac{kJ}{kg \cdot K}$

Heat capacity $C\_st\_D := m\_st\_D \cdot c\_st\_D = 2371.154 \cdot \frac{kW}{K}$

Flue gas

Inlet enthalpy $h\_fg\_in := 1314.7 \frac{kJ}{kg}$

Exit enthalpy $h\_fg\_ex := h\_fg\_in - \frac{Q}{m\_fg\_D} = 1196.142 \cdot \frac{kJ}{kg}$

Average temperature $T\_fg\_D := \frac{T\_fg\_in + T\_fg\_ex}{2}$

Conductivity $k\_fg\_D = 0.071\,m \cdot \frac{W}{m^2 \cdot K}$

Viscosity $\mu\_fg\_D = 45.278 \times 10^{-6} \frac{kg}{m \cdot s}$

Specific heat $c\_fg\_D := \frac{h\_fg\_ex - h\_fg\_in}{T\_fg\_ex - T\_fg\_in} = 1.153 \cdot \frac{kJ}{kg \cdot K}$

Heat capacity $C\_fg\_D := m\_fg\_D \cdot c\_fg\_D = 922.615 \cdot \frac{kW}{K}$

Heat transfer calculations

Minimum heat capacity $Cmin := min(C\_st\_D, C\_fg\_D) = 922.615 \cdot \frac{kW}{K}$

Maximum heat capacity $Cmax := max(C\_st\_D, C\_fg\_D) = 2371.154 \cdot \frac{kW}{K}$

Heat capacity ratio $Rc := \frac{Cmin}{Cmax}$

Heat exchanger effectiveness $eff\_D := \frac{Q}{Cmin \cdot (T\_fg\_in - T\_st\_in)} = 0.228$

NTU calculation $NTU :=$

$count = 0$

$NTU1 = 0$

$NTUx = 0.01$

while $|NTU1 - NTUx| > 0.001 \wedge count < 20$

$count = count + 1$

$NTU1 = NTUx$

$eff1 = 1 - e^{\frac{NTU1^{0.22}}{Rc} \cdot \left(e^{-Rc \cdot NTU1^{0.78}} - 1\right)} - eff\_D$

$NTU2 = 1.001 \cdot NTU1$

$eff2 = 1 - e^{\frac{NTU2^{0.22}}{Rc} \cdot \left(e^{-Rc \cdot NTU2^{0.78}} - 1\right)} - eff\_D$

$NTUx = \frac{NTU1 \cdot eff2 - NTU2 \cdot eff1}{eff2 - eff1}$

$NTUx$

Number of transfer units $NTU = 0.278$

Overall heat transfer coefficient

$$UA\_tot := NTU \cdot Cmin = 0.257 \cdot \frac{MW}{K}$$

Convective heat transfer coefficient

$$UA\_cD := UA\_tot - Rr \cdot \frac{Q}{T\_fg\_D - T\_st\_D} = 0.207 \cdot \frac{MW}{K}$$

σεAF term

$$\sigma\varepsilon AF := Rr \cdot \frac{Q}{T\_fg\_D^4 - T\_st\_D^4} = 17.426 \times 10^{-6} \frac{kg \cdot m^2}{K^4 \cdot s^3}$$

Operating calculations

Average steam temperature

$$T\_st := \frac{T\_st\_in\_act + T\_st\_ex\_act}{2}$$

Average flue gas temperature

$$T\_fg := \frac{T\_fg\_in\_act + T\_fg\_ex\_act}{2}$$

Conductivity

$$k\_fg = 0.065\, m \cdot \frac{W}{m^2 \cdot K}$$

Viscosity

$$\mu\_fg = 41.717 \times 10^{-6} \frac{kg}{m \cdot s}$$

Specific heat

$$c\_fg := 1130 \frac{J}{kg \cdot K}$$

Convection heat transfer coefficient

$$UA\_c := UA\_cD \cdot \left(\frac{m\_fg}{m\_fg\_D}\right)^{0.6} \cdot \left(\frac{k\_fg}{k\_fg\_D}\right)^{\frac{2}{3}} \cdot \left(\frac{c\_fg}{c\_fg\_D}\right)^{\frac{1}{3}} \cdot \left(\frac{\mu\_fg}{\mu\_fg\_D}\right)^{\frac{1}{3}-0.6} = 0.212 \cdot \frac{MW}{K}$$

Radiation heat transfer rate

$$Qr := \sigma\varepsilon AF \cdot \left(T\_fg^4 - T\_st^4\right) = 9.438 \cdot MW$$

Radiation heat transfer coefficient

$$UA\_r := \frac{Qr}{T\_fg - T\_st} = 0.04 \cdot \frac{MW}{K}$$

Overall heat transfer coefficient

$$UA := \left(\frac{1}{UA\_c + UA\_r}\right)^{-1} = 252048.1 \cdot \frac{W}{K}$$

Appendix E. Reciprocal analysis results

The reciprocal analysis results for evaluating the linear dependency of conduction resistance fault vectors with the different measurement combinations in Table 4-1 are shown Figure 8-7.

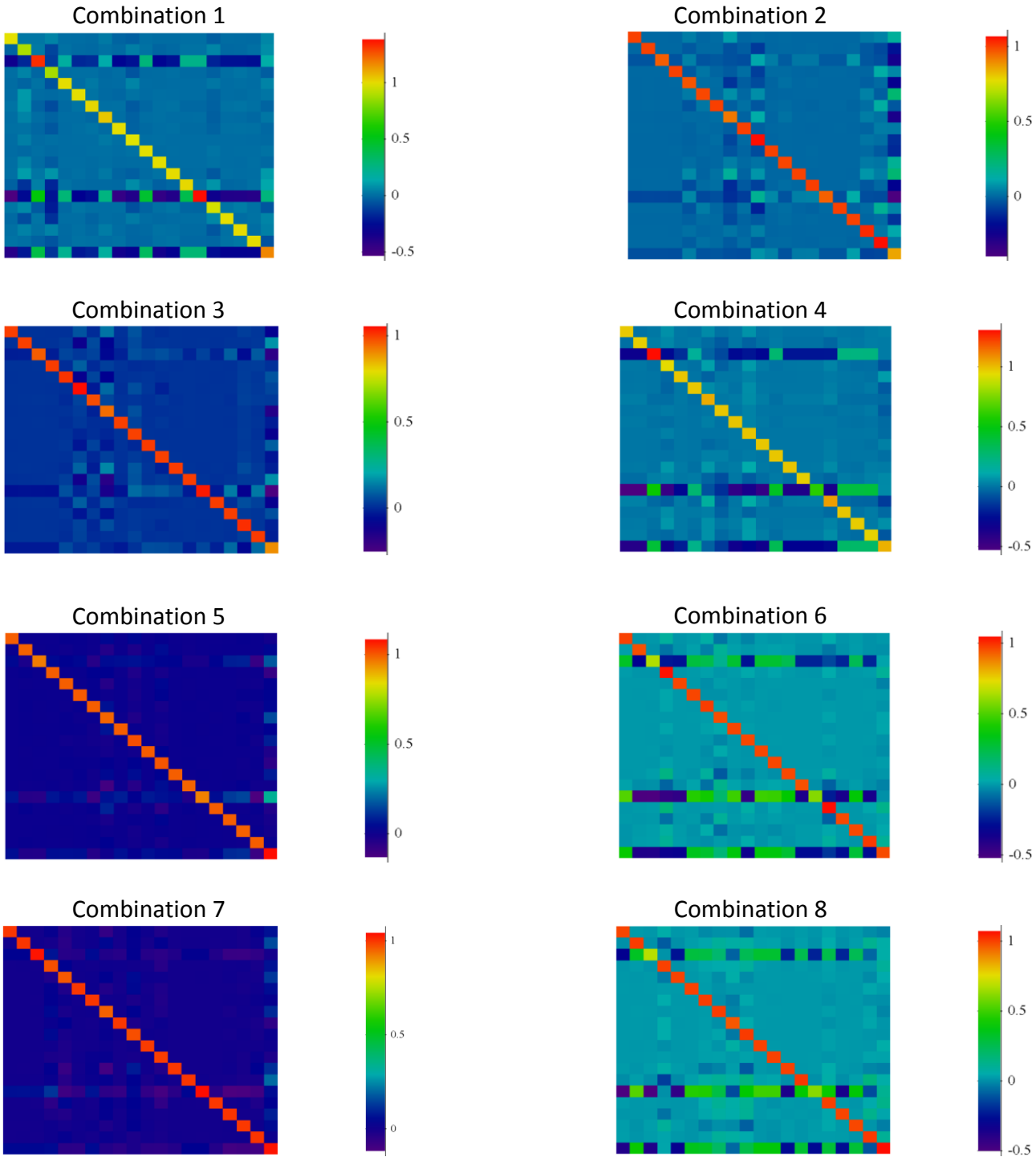
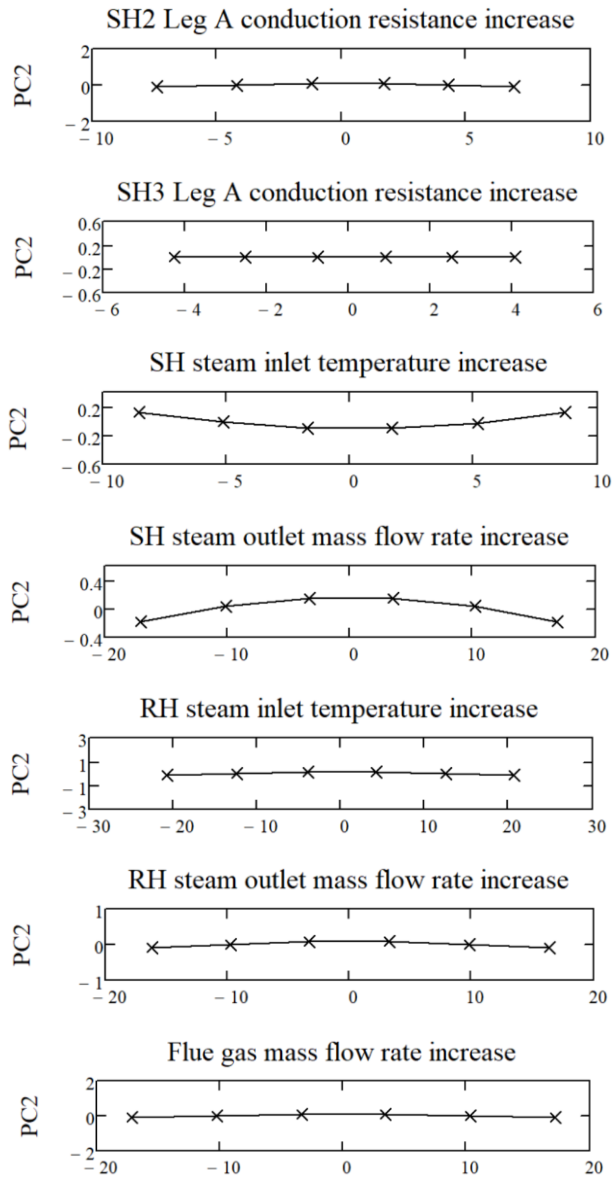


Figure 8-7: Reciprocal analysis results for the conduction resistance fault matrix with different measurement combinations

Appendix F. Fault vector propagation in PCA space

The major directions of fault vector propagation by increasing fault magnitude are reduced onto two dimensions using PCA. This is plotted on equal scales in Figure 8-8 for process boundary condition and conduction resistance faults not included in Figure 4-17. Only one leg per heat exchanger stage is shown for the conduction resistance faults as the results of the other legs are similar.



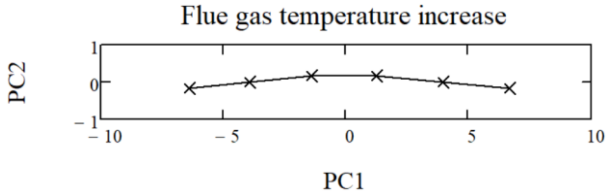


Figure 8-8: Approximating multidimensional fault vector movement with two dimensions